DOCUMENT RESUME

ED 290 760 TM 011 000

AUTHOR	Arter, Judith A., Ed.; Salmon, Jennifer R., Ed.
TITLE	Assessing Higher Order Thinking Skills: Issues and Practices Conference Proceedings (Clackamas, Oregon, October 1-2, 1987).
INSTITUTION	Northwest Regional Educational Lab., Portland, OR. Center for Performance Assessment.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Nov 87
CONTRACT	400-86-0006
NOTE	186p.; Parts of document contain small print.
PUB TYPE	Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	*Cognitive Measurement; Cognitive Tests; Creativity Tests; *Critical Thinking; Elementary Secondary Education; *Evaluation Methods; Item Banks; Multiple Choice Tests; *Test Construction
IDENTIFIERS	*Higher Order Skills; *Thinking Skills

ABSTRACT

The following discussions and workshops were presented at a conference on assessing higher order thinking skills organized by the Northwest Regional Educational Laboratory: (1) "Assessing Higher Order Thinking Skills--Issues and Practices" by S. Rankin; (2) "Writing Multiple-Choice Critical Thinking Items (part 1)" by S. P. Norris; (3) "Writing Multiple-Choice Critical Thinking Items (part 2)" by S. P. Norris; (4) "In-classroom, Informal Assessment of Students' Thinking Skills" by K. Bumgarner; (5) "Sow's Ears into Silk Purses--How To Take Unlikely Looking Textbook Material and Make It Testable for Critical Thinking" by C. Missimer; (6) "How To Select a Test of Higher Order Thinking Skills" by J. A. Arter; (7) "Diagnosing Thinking Deficiencies in the Classroom" by S. Wasserman; (8) "Assessing Creativity" by A. Halstead; (9) "Assessing Higher Order Thinking Skills in Hawaii" by J. A. Arter and L. Paule; (10) "Test Publisher's Panel" by C. Ross, G. H. Roid, M. Hill, and K. Bumgarner;(11) "Higher Order Thinking Skills and Item Banking" by R. Naccarato and R. K. Miller; (12) "Alternative Formats for Assessing Higher Order Thinking Skills" by E. S. Quellmalz; (13) "Classroom Assessment of Higher Order Thinking Skills" by R. Stiggins and E. S. Quellmalz; and (14) "Assessment Strategies to High Order Thinking Skills" by C. Ross." Appendices include the conference schedule and addresses of the presenters and conference participants. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Assessing Higher Order Thinking Skills

Issues and Practices

CONFERENCE PROCEEDINGS

Editors
Judith A. Arter
Jennifer R. Salmon

November 1987

Sponsored by

Northwest Regional Educational Laboratory
ESD #112, Vancouver, Washington
Oregon Department of Education
Northwest Evaluation Association
Idaho Association of School Administrators
Office of the Superintendent of Public Instruction, Washington

November 1987

These proceedings are published by the Test Center of the Northwest Regional Educational Laboratory, a private nonprofit corporation. The work contained herein has been developed under Contract 400-86-0006 for the Office of Educational Research and Improvement (OERI), U.S. Education Department. The opinions expressed in this guide do not necessarily reflect the views of OERI and no official endorsement by the office should be inferred.

TABLE OF CONTENTS

Page

Introduction 1

Session Summaries

Keynote Address *Assessing Higher Order Thinking Skills - Issues and Practices* by Dr. Stuart Rankin 3
Handout - *The Rankin-Hughes Framework of Thinking Skills* 7

Writing Multiple-Choice Critical Thinking Items (part 1) by Stephen P. Norris 17
Handout - *Some Guidelines for Writing Multiple-Choice Critical Thinking Items* 19

Writing Multiple-Choice Critical Thinking Items (part 2) by Stephen P. Norris 23
Handouts - *Principles for Appraising Observations* 25
Annotated List of Critical Thinking Tests 26
References on Critical Thinking Testing 28
A Traffic Accident 29

In-classroom, Informal Assessment of Students' Thinking Skills by Kenneth Bumgarner 31
Handout - *Thinking Log* 33

Sow's Ears Into Silk Purses - How to Take Unlikely-Looking Textbook Material and Make It Testable for Critical Thinking by Connie Missimer 41
Handout - *Test of Critical Thinking Skills* 43

How to Select a Test of Higher Order Thinking Skills by Judith A. Arter 59
Handout - *Higher Order Thinking Skills Tests* 63

Diagnosing Thinking Deficiencies in the Classroom by Selma Wasserman 73
Handout - *Assessing Thinking-Related Behavior* 74
Teacher Rating Instrument 86

Assessing Creativity by Anita Halstead 101

Assessing Higher Order Thinking Skills in Hawaii by Judith A. Arter and Lynde Paule 103
Handout - *FPO III Content Description* 105

Test Publishers' Panel
Catherine Ross, representing Academic Therapy,
Gale H. Roid, representing Psychological Corporation
Margaret Hill, representing SRA
Moderator: Ken Bumgarner 111

Higher Order Thinking Skills and Item Banking by Richard Naccarato and Ray K. Miller 113
Handouts - *Item Bank Data Sheet* 114
Item Bank Design 118
Item Retrieval Characteristics 125
Resource Bibliography 126
Taxonomy Cross-Classification 127
NWEA Item Bank Information 128

Introduction

This document reports the proceedings of a conference on assessing high order thinking skills organized by Northwest Regional Educational Laboratory and cosponsored by:

- Northwest Regional Educational Laboratory (NWREL)
- Oregon Department of Education
- Office of the Superintendent of Public Instruction, Washington
- Northwest Evaluation Association (NWEA)
- ESD 112 in Vancouver, Washington
- Idaho Association of School Administrators

The conference was held October 1 and 2, 1987 in Clackamas Oregon. The program for the conference is reproduced in Appendix A. The list of participants is provided in Appendix B. The purpose of the conference is best summed up by Gary Estes, Director of Evaluation & Assessment at NWREL, program introduction.

> The Northwest Regional Educational Laboratory, in conjunction with five regional associations and departments of education, is pleased to welcome you to two days of presentations and workshops on assessing higher order thinking skills (HOTS). Good assessment is essential if we are going to make decisions about the instructional needs and progress of students. One goal of this conference is to assist people who are deciding how to assess this critical area in our schools to make more informed decisions and to produce higher quality assessments. Another goal of this conference is to bring together people interested in the assessment aspects of thinking skills. One outcome might be a cooperative effort in item banking.
>
> The conference has been divided into two parts: Issues and Practices. The afternoon of the first day is designed to elicit perspectives about assessing higher order thinking skills. Issues such as how should HOTS be assessed, the quality of existing tests, how do we define HOTS and what will the future bring will be discussed. This part will set the stage for the second day's activities.
>
> The second day emphasizes practices. There are three strands: daily, less formal classroom assessment; formal assessment procedures that may be used at the district level; and training on how to write test questions which assess HOTS.
>
> The presenters and participants in the eighteen sessions will make each of the sessions a worthwhile exploration of the issues and practices for assessing higher order thinking skills.
>
> Once again, welcome! We are looking forward to a productive conference

These proceedings were produced by persons assigned to each session to take notes on the essential points covered by each session. All presenter handouts are included.

One outcome of this conference might be a cooperative NWEA-type project in the area of higher order thinking skills. NWEA is a consortium of school districts which work together to solve problems of mutual interest. Projects in the past have been in the areas of basic skills (resulting in good-quality item banks in the areas of reading, math and language arts) and science (which currently has addressed curriculum issues and has produced 7,000 test questions). The exact nature of the cooperative effort is developed by those districts, state departments of education and other organizations which participate. We invite readers who may be interested in such an effort to contact NWREL (503-275-9500) or NWEA (206-839-3932).

Assessing Higher Order Thinking Skills -- Issues and Practices

Keynote Address

Stuart Rankin

Dr. Rankin's objective for his address was to clarify assessment issues surrounding measuring higher order thinking skills (HOTS). He did not address the need to teach HOTS because he assumed that the audience recognized the need for this. Given that we need to teach HOTS, then it follows that we must assess students so that we can plan the proper instruction.

Dr. Rankin asserted tha. most of the major issues surrounding HOTS assessment are issues of validity -- how do we obtain a valid view of how students think? Dr. Rankin used his "Rankin-Hughes Framework Of Thinking Skills" to illustrate the following assessment issues surrounding HOTS:

1. **The content to be covered by the assessment should be based on some theoretical framework.**

We should have a rationale for what we teach and what we assess. This rationale is best established by using a good framework of HOTS skills. Dr. Rankin briefly discussed Bloom's taxonomy and the framework surrounding the Philosophy For Children Program. Then he used his own framework as an illustration of a reasoned and comprehensive framework. (The attached paper entitled "The Rankin-Hughes Framework of Thinking Skills" discusses this framework in more detail.)

2. **We should consider the full-range of HOTS skills in our instruction and assessment, not just the easiest ones to teach or assess.**

Dr. Rankin asserted that many current tests (especially those in paper and pencil format) test a narrow range of HOTS skills; namely, those that are amenable to assessment in that format. He once again used his framework to illustrate that thinking skills consist of many more areas than logical syllogisms, identifying assumptions and making inferences.

For example, we should not ignore such things as thinking processes such as concept formation, oral communication, problem solving and decision making; metacognition; creativity; critical thinking; and the affective side of HOTS.

3. **We should be very careful what we assess because assessment can drive the curriculum.**

Dr. Rankin sees thinking as a very complex task. We have a pool of skills, abilities, dispositions and processes, to be drawn from as needed. We should not view HOTS as being a fixed set of skills that are applied the same way for each use. He has a holistic view of how we solve problems -- we draw what we need as we need it. Skills should not be taught or assessed in isolation except to sharpen them. Therefore we should try to assess the thinking process as a whole rather than as separate skills. If we view thinking as separate skills, and worse yet, separate skills as listed by some fixed taxonomy, then we threaten to drive the curriculum in that direction.

A related issue is that any assessment is a sample of skills from a domain. We cannot assess the entire domain. The curriculum becomes restricted if we treat the content of the test as the entire domain of interest.

4. The purpose of assessment should be to plan instruction not to screen students.

Dr. Rankin feels that using tests of HOTS to screen students is nonproductive. If we use assessment devices at all, it should be to assist teachers to know how to help students.

5. We should spend more time developing test specifications.

Test specifications provide a blueprint for a test -- what is to be covered, how each skill will be assessed, and how may questions or exercises will be written to measure each skill. Dr. Rankin asserts that test specifications, in addition to helping develop a good test, can also be useful instructionally. Test specifications should clearly define each skill, should provide a sample question or exercise that illustrates how to measure that skill, criteria for judging whether the student has demonstrated the skill to be measured, and what the implications for instruction are if the student does not demonstrate the skill. In addition, these test specifications should cover all areas of importance not just those easy to measure.

In order to develop good test specifications (and in the process, good instructional aids) there needs to be consultation between curriculum people and test developers. The curriculum people need to specify what should be measured (and taught). Then the groups can work together to figure out how to measure it adequately.

6. We should try to control for prior knowledge when development assessments of HOTS.

The issue here is that student knowledge can alter a task so that instead of measuring thinking, it really measures something else such as recall of knowledge or vocabulary. Dr. Rankin pointed out that kids with good vocabularies do well on thinking skills tests. Vocabulary level can contaminate the results.

7. We need to consider assessing skills in more formats than multiple-choice.

If we really want to gather information about most of the more interesting thinking skills, we need to consider other assessment formats than multiple-choice. Other formats include short answer, essay and performance tasks.

8. Our notions of what thinking skills are should not be fixed.

We are in a continual process of better defining what thinking is and what thinking skills are. As we reconceptualize this area we should be redefining our taxonomies, teaching strategies and assessment activities. If our skills lists become too fixed, we are threatened by rote approaches. We need to be continually thinking about thinking. We may be better able to define thinking by developing mathematical models of the thinking process. Dr. Rankin invited the audience to pursue this direction.

9. We should disaggregate information about the HOTS performance of students.

Dr. Rankin feels that we should use the information we collect about students to improve instruction. One way to do this is to look at the relative performance of various groups of students. We should confront head-on issues such as difference in performance between ethnic groups.

10. Test developers should be critical thinkers too.

We need to be constantly thinking about what we are doing and why. We need to consider alternatives, be creative and do what makes sense.

Panel Response

Dr. Stephen Norris, Dr. Edys Quellmalz and Dr. Ken Bumgarner responded to Dr. Rankin's comments.

Dr. Norris indicated that he agreed with Dr. Rankin on the following points:

1. Evaluation must be based upon a comprehensive theoretical framework of good thinking. This framework must have the flexibility to evolve.

2. We need to conceive of good thinking as selecting from a pool of skills those which are appropriate for a problem rather than as rigidly applying a sequenced set of skills.

3. We need to focus our attention on contexts of reasoning and on the complexity of coordinating many skills in the solution of problems in content, rather than on the use of decontextualized, isolated skills.

4. More attention needs to be paid to the attitudes and dispositions associated with good thinking.

5. We must remember that critical and creative thinking do not mark off completely separate acts. They overlap and are mutually supportive.

6. Be careful that your thinking tests do not merely test for background knowledge.

7. Teachers and test developers need to be good thinkers themselves.

Dr. Norris emphasized that choice of theoretical frameworks is very important; some are better than others. In particular, use of Bloom's Taxonomy to conceptualize good thinking (a use for which Bloom never intended it) can lead to serious problems:

1. By "knowledge" Bloom means mere recall. In its everyday use and use in much educational theorizing, knowledge means far more than this.

2. Comprehension is near the bottom of Bloom's taxonomy, but comprehension can be one of the most complex tasks in which human beings engage.

3. The categories of the taxonomy are too vague to provide much guidance for testing; for example, the Evaluation category provides no criteria for making evaluations, students' knowledge of which should be tested.

4. In general, the notion "Higher Level Thinking" is dangerous. Many so-called low level skills (e.g., reading and observation) are very complex when done properly. Much depends on the context.

Finally, Dr. Norris commented that whereas Dr. Rankin had claimed that critical thinking is not a bunch of skills, it is an attitude, critical thinking is best thought of as a combination of both skills and attitudes.

Dr. Quellmalz expanded on several of Dr. Rankin's points. With respect to theoretical frameworks, she feels that there are currently four to five good, comprehensive theoretical frameworks. We need criteria for evaluating them. We also need to look to other fields, such as cognitive psychology to refine frameworks. Second, Dr. Quellmalz emphasized the need to assess thinking skills within some meaningful context -- purposeful, sustained efforts. Third, she believes that some skills are general and can be applied across content domains.

Fourth, she agreed with Dr. Rankin that we need to broaden our assessment procedures to include other formats than multiple-choice. She suggested portfolios of student work, interviews, and learning logs.

Dr. Bumgarner agreed with Dr. Rankin that we should use assessment for instructional planning; but, it also promotes what is of value. Therefore, it is important to assess it so that it is taught. Dr. Bumgarner feels that we should stress all thinking skills not just "higher order" ones. Recall of knowledge is important too. He also believes that there is an increasing gap between the general state of information on teaching and assessing HOTS and common classroom practice.

Dr. Bumgarner added to the list of issues presented by Dr. Rankin. We need to consider the effect of testing on students, both in terms of self-concept and in terms of what the types of tests we give to students tell students about what is important. An example of the latter point is that if we give students tests having only one right answer, they may come to believe that everything has a right answer, an unrealistic expectation of the real world.

Finally Dr. Bumgarner stressed that we are currently too dependent on commercially developed tests. This, as Dr. Rankin pointed out, leads to reductionism. We need to continue to develop in teachers the ability to be valid assessors because the on-going interaction of teachers and students should be the arena for continual assessment and instructional planning. Finally we need to establish public faith in teachers' ability to assess learning.

THE RANKIN - HUGHES FRAMEWORK* OF THINKING SKILLS

The major thinking processes have much in common. Problem solving, decision making, disciplined inquiry, concept formation, principle formation, comprehending, and composing generally begin with an unresolved problem, disequilibrium, or indeterminate situation and move to a resolution, new equilibrium, or new meaning.

In one sense the common features of the processes can be derived by examining the steps in each of the major processes and noting their similarities. In another sense the common process is generic to, imbedded in, and indeed drives the major thinking processes.

Rankin and Hughes have designed a framework which spells out steps or stages that appear to be common to all of the major thinking processes. The framework also shows a way the individual thinking skills can be related to the processes. Some skills are more useful at certain steps in a process than at others. The framework, however, is not intended to restrict the place in the process where a skill may be used or to imply that each process always follows the same ordered sequence of skills.

This framework does help to reinforce the idea that thinking skills are part of a larger process and that, although they can be described and sometimes taught in isolation, they are best viewed as tools to be sharpened and used in constructing meaning, solving problems, and creating products and ideas.

It should be noted that thinking skills are often imbedded in each other. For example, comparing and classifying are skills that may be used as part of recognizing patterns and relationships or of inferring. Thinking skills are alike in that they usually involve mental acts which combine two or more representations of elements. The elements represented may be objects, units of experience, or observations; they may be attributes, qualities or characteristics; they may be symbols, concepts, or principles; they may be simple or complex; and they may be single items or patterns or relationships.

The generic process presented in this framework has seven steps or stages. They are (1) focus, (2) gather information, (3) organize information, (4) analyze information, (5) generate ideas, (6) synthesize, (7) evaluate and apply. Some thinking processes will involve all of the stages, while other thinking processes only involve some of the steps depending on how complex the goal attainment becomes.

---

*Designed by Stuart C. Rankin, Deputy Superintendent, Detroit Public Schools, and Carolyn S. Hughes, Assistant Superintendent, Oklahoma City Public Schools for the Association for Supervision and Curriculum Development, 1986.

Outline of Framework

1. Focus

   - Sense problem
   - Define problem
   - Set goal

2. Gather Information

   - Observe
   - Recall
   - Question

3. Organize Information

   - Represent
   - Compare
   - Classify
   - Order

4. Analyze Information

   - Distinguish and clarify components and attributes
   - Determine accuracy and adequacy of arguments
   - Recognize patterns and relationships
   - Identify central element

5. Generate Ideas

   - Infer
   - Anticipate
   - Discover relevant outside structures
   - Restructure

6. Synthesize

   - Summarize
   - Integrate
   - Develop outcome

7. Evaluate and Apply

   - Establish criteria and standards
   - Verify
   - Revise
   - Transfer

1. Focus

The first step in a thinking process is focusing attention. When the mature thinker senses a problem, he or she will often define the problem and set a goal before proceeding to gather or to process information. When the problem is less complex, or when the thinker is less mature, problem sensing and defining may be unconscious steps which lead the thinker immediately to focus on a goal and begin to gather information.

Sense Problem

The skill of problem sensing is the ability to discern a lack of fit between what is and what is needed; a growing awareness of discrepancy, disequilibrium or need; or a consciousness that something is missing or needed to restore or to create a meaningful whole. This skill involves recognition of a need to know or to understand. It may include awareness of a desire to produce or to experience something. Developing the skill of problem sensing is based on attitudinal components of curiosity or concern in combination with knowing the factors critical to a particular situation. The situation for problem sensing may be as simple as awareness of a fuel indicator approaching "empty" or as complex as recognizing when a politician's words and actions are incongruent.

Define Problem

As an individual moves from consciousness of a need, problem, or discrepancy to engagement with it, he or she focuses attention on the situation. Direct instruction and practice can result in growing ability to focus and control attention. Identification of examples or relevant variables helps to bring a problem into focus. Such focus enables the thinker to determine whether a problem does indeed exist, and to confront the problem and to perceive it clearly. Developing skill in defining a problem includes determining assumptions and making an initial estimate of success criteria so that one will know when the problem or discrepancy is resolved. Problem definition may be stimulated by such a question as "What is it that leads you to wonder whether you can trust what that politician says?"

Set Goal

Once an individual has become conscious of a need, problem or discrepancy and confronted various aspects of the situation, a goal is needed to guide and justify effort. This is a commitment of energy which involves determining direction and identifying possible outcomes. Goal setting involves anticipating types of information and resources needed, identifying processes to use, and choosing a tentative medium of expression. As students grow in metacognitive awareness, goal setting will include planning a thinking strategy: the combination of thinking skills needed to achieve the purpose. Determining direction may be guided by a question such as, "How could we determine whether that politician's actions are consistent with what he says in this speeches?"

2. Gather Information

The foundation skills for thinking are those which make available to the mind what one thinks about. Information gathering skills provide the basic substance or content which becomes the raw material for cognitive processing. The skills which make that substance available to the thinker are observing to gather first-hand sensory input, recalling to activate prior knowledge, and questioning to seek out new information.

Observe

The ability to focus attention and perception is a primary element in the skill of observing -- the use of the senses to gain direct, first-hand information. Sustained focus and systematic search add power to the skill of observing. When words are used to describe and differentiate sensory experiences, awareness grows helping students become better observers and better reporters of their observations. When observing, the ability to note unfamiliar characteristics or terms can provide valuable clues for further thinking. The skill of observing is guided by asking such questions as, "What do you notice about _____?"

Recall

Recalling is the skill of accessing or activating prior knowledge from memory. It may be an active, thoughtful process of reconstruction -- systematically recounting specific information or events from prior experience, or it may be simple association such as recalling a name or number. The content for recalling may be experiences, information, concepts, structures or categories.

Question

While observing and recalling provide information which is immediately available to the senses or within one's previous knowledge or experience, questioning is a primary skill for gathering new information. Questioning is guided by awareness of the need for relevant, accurate, specific information. The clearer one's understanding of a problem or issue, the more focused will be the questions one uses to elicit information. Skillful questioning, based on what one needs to know, enables one to gain optimum value from reading, listening, interviewing, or discussion.

3. Organize Information

The limited number of isolated bits of information which one can keep in mind at a time leads to the need to organize information. Another reason for organizing information is to develop new insights and understandings. Representing is an organizing skill which enhances ones ability to communicate and to retrieve information when it is needed. The organizing skills of comparing, classifying, and ordering build a foundation for information analysis.

Represent

An early step in processing information which has been gathered is translating it into some form which is useful or appropriate to the task. Representing is changing the form but not the substance or value of the information. A name or label may be assigned to represent it. Outlines, graphs, tables, drawings, models or mental pictures may be used to represent the information. The thinker may represent information using logical, mathematical, musical or visual symbols. Intentional mental encoding of information so that it can be easily retrieved may include multi-sensory recall activation also known as deep processing or the use of memory frameworks such as pegwords, familiar place clues, number/word clues, or other mnemonic devices.

Compare

One compares when one identifies likenesses or similarities. Contrasting is identifying differences. Developing proficiency in comparing and contrasting is enhanced by teaching them separately. In this framework the skill of comparing has been expanded to include contrasting as well as comparing. This basic skill for organizing information forms the foundation for developing concepts and for classifying information as well as for the thinking skills needed to analyze, and to generate ideas.

Classify

Classifying is identifying examples of a concept by ensuring that all concept attributes are present. The skill of classifying requires the knowledge (or invention) of a definition or concept appropriate for the information at hand. A component skill for classifying is grouping. Grouping is putting together items which one perceives to be related by some common characteristic(s). Grouping is based on comparing and may lead to classifying if a concept label or name is given to the group.

Order

One orders by first establishing a scheme or criterion for sequencing information. The criterion may be based on such physical factors as position, size, intensity, or duration, or it may be based on abstract qualities such as preference, priority or agreement. Once the scheme or criterion is determined, it is used to place information in a sequence.

4. Analyze Information

Analysis requires a look inside the components to distinguish among them, clarify them, determine their adequacy, or rearrange them. It involves the recognition of relationships and the identification of central elements. Examination is made of parts and patterns in order to understand them and to suggest possible solutions. The skills of analyzing information are central to critical thinking.

Distinguish and Clarify Components and Attributes

The functions of different elements in a situation must be made clear. Distinctions can be made between facts and opinions, between inferences and assumptions, and among findings, conclusions, and implications. Attributes and meanings may require identification and clarification.

Determine Accuracy and Adequacy of Arguments

A fundamental step of analysis is establishing accuracy. Some elements may be biased, false, or ambiguous. Others may be inconsistent, irrelevant, or contradictory. Claims may be unwarranted. Omissons may be noted which limit comprehensiveness or break the chain of an argument. The accuracy of facts and the reliability of sources must be determined.

Recognize Patterns and Relationships

Patterns can be seen among elements whether those elements are concepts, objects, principles, attributes, qualities, or actions. Among the components various relationships may be identified which are temporal or spatial, sequential or correlational, or transitive or syllogistic. Casual or hierarchical relationships may be found. Analysis requires the recognition of these patterns. Comprehension and recall are enhanced when readers can recognize frequently used paragraph patterns such as comparison, classification, sequence, or causation. The study of number patterns and geometric patterns enhances the understanding of mathematics.

Identify Central Element

Many situations have a central theme or main idea. The essence of something may be described as its kernal, or key element, or critical incident. It may be a central fibre, a musical theme, a color or geometric pattern, of some other unifying element that adds meaning or impact to the whole.

5. Generate Ideas

Where analysis looks inside a situation, generation of new ideas requires one to extend and expand. Extension projects beyond the situation, and expansion makes distant connections that may lead to a whole recasting of the problem situation with additional components into a new structure. Extending depends upon inference and anticipation, while expanding makes use of invention, discovery, imagination, and creativity.

Infer

New meanings, relationships, and characteristics can be implied by or derived from a given situation. Generalization, conclusions, inductions, and deductions lead to the new understandings. New meanings of a passage are inferred by extending the meanings contained in and supported by the passage. A casual relationship may be hypothesized about two events which are known to

have occurred but the relationship is not given; previous experiences may suggest the relationship. Hypotheses about the characteristics or attributes of a component or subject may be supported by relating clues in the situation to previous knowledge, to experiences, or to the opinions of authorities.

Anticipate

Trends can often be projected. Possible results or occurrences can be identified. Probabilities may be determined for those possible consequences. Estimates of the magnitude or strength of a relationship can be made based on the situation and on past experiences. Predictions and forecasts can be made beyond on estimates and probabilities. Though anticipating includes the skill of inferring, it projects further beyond the known and requires identification of necessary or sufficient conditions.

Discover Relevant Outside Structures

When a problem cannot be resolved by analysis or extension, or when a more creative solution is ought, it is necessary to look beyond the elements of the problem situation. One searches in past experiences and present observations which are external to the problem situation in order to discover or invent a connection (isomorphism) with the problem situation. This type of thought usually occurs following withdrawal of the focus of attention on the problem situation. It may require the restraining of impulsivity or the withholding of judgment. Robert Penn Warren once called it, "Stop scratching in order to build up the itch." The lessening of focus appears to let the mind be free to roam about to metaphorical situations taking advantage of imagination and creativity to find unexpected new suggestions that can lead to solutions. These new connections are isomorphic to some part of the problem solution in that there is a correspondence of some elements and some underlying structure.

Restructure

Further examination of the relevant outside structure in respect to the problem situation can produce analogies, metaphors, and other models which in turn may furnish hypotheses about possible solutions. This use of outside models serves to restructure or recast the indeterminate or unresolved problem situation or disequilibrium into a new structure which can form the basis of a solution. This restructuring may be quite creative and is sometimes called a paradigm shift. Sometimes it is not just the elements of the problem situation that are recast, but the description or understanding of the problem itself that is restructured.

6. Synthesize

The results of analyzing, and generating ideas are new themes, meanings, structures, and design features which need to be pulled together to create a new understanding, product, or solution. Summaries are made of the key components; the components are integrated into a meaningful design; and the actual outcome is developed.

Summarize

The components, findings, or design features which were discovered or created during analysis, and generating ideas are stated or represented in a condensed, concise, and comprehensive manner. The main metaphor for the poem is identified; the key elements of the decision are determined; the parts of the solution are named; the colors and the materials are chosen.

Integrate

To integrate is to put the parts together in a whole. As does summarizing, this skill requires identification of the parts, but it also requires the organization of the parts so that their interrelationships and structures form a meaningful whole. Complete integration goes beyond internal organization of the parts and requires that the new whole fits properly into its external world as well. The outline is written; the sketch is made, the clay model is formed; the dance movements are put in order.

Develop Outcome

When an integrated designed of a solution is formed, the synthesis is not finished until the full development of the outcome is accomplished. This skill varies with the medium and form of the outcome, but in each case it requires selection, execution, and completion of the details of the product. All refinements of the meaning are made; the picture is painted; the decision is made; the concept is formed.

7. Evaluate and Apply

A full goal-oriented thinking process is not complete until the outcome is evaluated and applied. Standards are set for internal and external validity, and the product is verified with respect to these standards. Revisions are made where necessary. When the criteria are met, the outcome is available for attaining the goal and for applying or transferring to other situations.

Establish Criteria and Standards

Two kinds of criteria must be set in order to judge the quality or adequacy of the solution, decision, meaning, or product. First, the outcome must have internal validity. It must be produced in the domain of the problem; it must resolve the disequilibrium that motivated the process; and it must be usable by the one who produced the outcome. Sometimes these standards for internal consistency are explicit, sometimes implicit. Second, the outcome must work in the real world. Standards for quality of an art work may be used. The finding or meaning may reinforce or refute other knowledge. The solution may have wide or narrow applicability. It may be easy or hard to replicate. Criteria are used to set standards against which the outcome is to be judged. Standards, to be most useful, should be measurable.

Verify

After the standards are set, they are used to judge completeness, quality, and range of applicability. Outcomes are verified by comparing them with standards to determine if the solution, meaning, or product is consistent with the conditions accepted when the goal was set. Assumptions and delimitations must be met. Also, the outcome must be of value, e.g., useful, beautiful, scarce, original. In addition it must resolve the initial problem situation. Both the internal and the external criteria must be met.

Revise

Evaluation results in decisions to accept, reject, or modify the whole outcome or its parts. If no revisions are needed, then the whole is accepted. Often some components (elements, structures, details) are retained, others are eliminated or replaced and still others are modified. Sometimes new components must be added. The determination of whether revisions are needed, of the reasons for the revisions, and of the areas where revisions are needed is an evaluation skill. The actual revision leads one back to earlier steps in the thinking process. Whether one returns to problem definition, goal setting, information gathering, analysis, extension, or expansion is determined by the evaluation decision to revise.

Transfer

When the outcome is finally accepted, that acceptance may be a new meaning which itself is the application of the thinking process. Often outcomes can be used repeatedly for other similar situations. One assimilates or accommodates the outcome and it becomes part of experience. If stored well in memory, it will be retrievable for application to like problem situations and transferable to some situations quite unlike the problem situation. Such transfer requires (1) experiencing the process, (2) storing in a retrievable fashion, and (3) using thinking skills from the generating ideas category to help bridge the gaps. A final application of any solution or outcome may be the identification of a new problem.

Writing Multiple-Choice Critical Thinking Items (part 1)

Stephen P. Norris

Dr. Norris suggested that while multiple choice tests of critical thinking have limitations, especially when used to provide information on specific information on specific individuals, they can when constructed properly provide valuable information on groups. Multiple choice items do not, however, test for the orchestration of thinking skills on complex problems nor test for critical thinking dispositions.

To frame the discussion about writing good critical thinking items, Dr. Norris distributed a handout, Some Guidelines for Writing Multiple-Choice Critical Thinking Test Items. The greatest part of this session followed the "Suggestions for Specific Aspects of Critical Thinking" section of the handout. Key points in this section were illustrated with examples from specific test items in published tests.

Some Guidelines for Writing Multiple-Choice Critical Thinking Test Items

from Stephen P. Norris
and Robert H. Ennis,
Evaluating Critical Thinking, Pacific Grove, CA:
Midwest. (In Press)

by
Stephen P. Norris
for
Assessing Higher Order Thinking Skills:
Issues and Practices
The Northwest Regional Educational Laboratory
October 1-2, 1987

General Suggestions

1. Start with a good conception of critical thinking.

2. Identify those aspects of critical thinking that are most important in your situation. Indicate relative importance with a weighting scheme.

3. Decide on the purpose of the test (testing for mastery, comparing students with others via norms, making placement decisions, diagnosis, pretest-posttest comparisons).

4. Decide whether you want to test for critical thinking in a particular school subject or in the context of general knowledge.

5. Provide adequate context.

6. Three or four alternative answers are preferable to two because of chances for guessing correctly. But if there is a request for justification, a two-choice set quite acceptable. (Now you do not have a strict multiple-choice test.)

7. Interview a range of students of the sort for whom the test is designed to find out how they interpret the questions and why they gave their answers. Revise test items that are revealed to unfairly help or hurt some students.

8. Remember that writing multiple-choice critical thinking items is an art.

9. Careful scrutiny of items and multiple revisions are usually needed. Do not expect to make a good test in one trial.

Rules for Item Writing

1. Construct each item with one and only one correct or best answer.

2. Avoid "none of the above" and "all of the above" as choices when examinees are to choose the best, rather than precisely correct, answer.

3. Use either a direct question or an incomplete statement as the item stem.

4. Write items in clear and simple language.

5. State the central problem of the item clearly and completely in the stem.

6. Include most of the reading in the stem.

7. Base each item on a single, central problem.

8. Construct options homogeneous in grammatical form.

9. Include in the stem any words that would otherwise need repeating in each option.

10. Emphasize negative words or words of exclusion (e.g., "not", "except") and avoid such words when possible.

11. Place options at the *end* of the item stem not in the middle of it.

12. Arrange the options in a logical order, if one exists.

13. Make all options plausible to examinees who do not know the correct or best answer.

14. Avoid unintended hints based on:
    a. grammatical consistency or inconsistency between the stem and the options,
    b. repetition of key words in the stem and keyed option, or
    c. rote or other verbal associations between key words in the stem and the keyed option.

15. Avoid hints based on the:
    a. unusual length of the keyed option,
    b. degree of qualification stated in the keyed option or use of terms such as "never" and "always" in the unkeyed options,
    c. lack of independence and mutual exclusivity of the options,
    d. frequency with which the keyed option is placed in a given option position, or
    e. pattern of the location of the keyed position.

16. Avoid hints from one item to another.

Suggestions for Specific Aspects of Critical Thinking

Judging the Credibility of Sources and Observations

1. Have a comprehensive and defensible set of criteria for judging credibility.

2. Give a situation to which credibility criteria can be applied.

3. Construct items so that only one criterion applies or most plausibly applies.

4. Ask for comparative, not absolute judgments. For example, ask which, if either, of the sources is more credible on a topic; or ask which, if either, of two conflicting observations is more credible.

5. Ask for a justification of the answer, if there is time for an evaluator to read it. Realize that now the test is not strictly multiple-choice.

Identifying Unstated Assumptions

1. Distinguish the conclusion, pejorative force, and basis senses of "assumption".

2. Provide an argument or an explanation in which unstated assumptions are made.

3. Ask for comparative, not absolute judgments. For example, give several choices and ask "Which is probably taken for granted, though not stated?"

4. Make sure no alternative is a conclusion that someone might draw from the stem.

5. One and only one of the alternatives should complete, or strongly contribute to completing, a deductive connection from the reasons to the conclusion, or from the explanatory material to the fact to be explained.

6. Ask for a justification of the answer, if there is time for an evaluator to read it.

Inducing and Judging Inductions

1. Provide a situation with a conclusion to be judged on the basis of its being the best explanation of the facts.

2. Ask for comparative not absolute judgments about the conclusion. For example, ask either:
   a. which of two conclusions, if either, is more justified; or
   b. for the direction of support of the evidence (counts for, counts against, counts neither for nor against) rather than whether the conclusion is true, probably true, probably false, or false, or provided with insufficient data to go either way.

3. Ask for a justification of the answer, if there is time for an evaluator to read it.

Deduction

1. Provide an argument.

2. Absolute judgment is all right here. For example you can ask whether the conclusion follows necessarily, or contradicts, or neither. Or you can ask, "Which, if any, of these follows necessarily from...?"

3. For more sophisticated students make sure that there are some items in which nothing follows necessarily.

Writing Multiple-Choice Critical Thinking Items (part 2)

Stephen P. Norris

Dr. Norris began by stating multiple-choice tests of thinking skills face a number of limitations. In such tests, examinees do not have to provide reasons for the answers they choose. If we do not know the reasons why examinees respond as they do, then we are not able to justifiably conclude whether their responses are due to good or poor thinking.

At the same time, multiple-choice tests have advantages. They can be scored relatively easily; they can test a large number of aspects of critical thinking in a short period of time; and they can thoroughly cover a single aspect of critical thinking much more efficiently than any other form of test.

Dr. Norris asserted that by applying specific procedures during the *design* of the tests which *build in* the sort of relation we desire between good and poor thinking and keyed and unkeyed answers, respectively, we can preserve the advantages of both.

This test building activity was demonstrated by discussion of development of the Test on Appraising Observations.by Norris and King (Handout 1). It was emphasized that this test concentrated on only one aspect of critical thinking, judging the credibility of observations.

The first step in designing a test is to define the *domain* of what is to be tested. In the example for this session, knowledge and use of a set of *principles* of observational appraisal were being evaluated. The Principles for Appraising Observations were included in the handout and used as a framework for participants as they looked at the development of specific test items.

Trial items were then written to test various principles. The problem is how to determine if the items are good? Dr. Norris stated that an item would be good if when examinees thought in accord with the Principles, they chose the keyed answers, and if when they did not think in accord with the Principles, they chose unkeyed answers.

To use Norris' method for determining if an item is good for a specific test, the developer should choose a sample from the entire population to be tested. (Such a sample usually cannot be random, but you should strive for reasonable representativeness.) Each examinee should be asked to read the items, choose an answer and then tell the developer all they were thinking in making their selection. Ideally, all probes should be open-ended. But reticent examinees can be asked more leading questions which seek justification or specific information. Everything examinees say should be tape-recorded.

Based on their responses, examinees are given two scores:

1. Performance Score:
   For each item, 0 = item answered wrong,
   1 = item answered right.
   For entire test, total number of right answers.

2. Thinking Score
   - assessed independently from answer chosen

- For each item, 0-3 according to specific criteria (this scale can vary from test-to-test)
- For entire test, total for each item.

These scores can then be compared across all subjects for each item and thus provide an indication of the level of evidence for the validity of each item. The development of Evidence Scores using the scores above was discussed. This process was illustrated with specific examples from the development of the Test on Appraising Observations.

To accumulate Evidence Scores for each item over a sample of students, Norris developed a "Thinking/Performance Index Score" for each item and this was explained.

Dr. Norris summarized by saying that this method was a combined qualitative-quantitative approach to the validity of a test. The hope is that in using this design methodology we can capitalize upon the strengths of multiple-choice tests and minimize their weaknesses.

Taken from: Test on Appraising Observations Manual. Stephen P. Norris & Ruth King. St. John's, Newfoundland: Institute for Educational Research and Development, Memorial University of Newfoundland. (Copyright, 1985)

NORRIS
HANDOUT

TABLE A

Principles for Appraising Observations

I. Observation statements tend to be more believable than inferences based upon them.

II. An observation statement tends to be believable to the extent that the **observer**

1. is functioning at a moderate level of emotional arousal.
2. is alert to the situation and gives his or her statement careful consideration.
3. has no conflict of interest.
4. is skilled at observing the sort of thing observed.
5. has a theoretical understanding of the thing observed.
6. has senses that function normally.
7. has a reputation for being honest and correct.
8. uses as precise a technique as is appropriate.
9. is skilled in the technique being used.
10. has no preconceived notions about the way the observation will turn out.
11. was not exposed, after the event, to further information relevant to describing it.
(If the observer was exposed to such information, the statement is believable to the extent that the exposure took place close to the time of the event described.)
12. is mature.

III. An observation statement tends to be believable to the extent that the **observation conditions**

1. provide a satisfactory medium of observation.
2. provide sufficient time for observation.
3. provide more than one opportunity to observe.
4. provide adequate instrumentation, if instrumentation is used.
(If instrumentation is used in gaining access, then the statement tends to be believable to the extent that the instrumentation
a. has suitable precision,
b. has a suitable range of application,
c. is of good quality,
d. works in a way that is well understood,
e. is in good working condition.)

IV. An observation statement tends to be believable to the extent that the **observation statement**

1. commits the speaker to holding a small number of things to be true.
2. is corroborated.
3. is no more precise than can be justified by the observation technique being used.
4. is made close to the time of observing.
5. is made by the person who did the observing.
6. is strongly believed to be corroboratable by the person making it.
7. does not conflict with other statements for which good reasons can be given.
8. is made in the same environment as the one in which the observation was made.
9. is not about an emotionally-loaded event.
10. is the first report of the event provided by the speaker.
11. is not given in response to a leading question.
12. does not report a recollection of something previously forgotten.
13. reports on salient features of an event.
(Features of an event are salient to the extent that they are extraordinary, colourful, novel, unusual, and interesting, and not salient to the extent that they are routine, commonplace and insignificant.)
14. is based upon a reliable record, if it is based upon a record.
(If an observation statement is based upon a record, then the statement tends to be believable to the extent that the record
a. was made close to the time of observing.
b. was made by the person who did the observing.
c. comes from a source having a good reputation for making correct records.)

Annotated List of Critical Thinking Tests

Prepared by

Stephen P. Norris

for

Assessing Higher Order Thinking Skills:
Issues and Practices

The Northwest Regional Educational Laboratory

October 1 and 2, 1987

In compiling this list I do not intend to give specific endorsement to any of the tests. The state of the art in critical thinking testing is rather more primitive than desirable. Thus, each listed test suffers deficiencies and some tests are worse than others. However, they are essentially all that is available.

For guidance on choosing a critical thinking test I recommend the items under "References on Critical Thinking Testing." More specifically, I suggest the following guidelines when choosing a critical thinking test:

1. Pay close attention to the directions, the test items, and the scoring guide.
2. Take the test yourself and satisfy yourself that the scoring guide is reasonable for the students you wish to test.
3. Ask yourself whether the test really tests for critical thinking.
4. Ask yourself whether the test covers those aspects of critical thinking that you wish to assess.
5. Read the test manual looking for evidence on the validity and reliability of the test.

The Cornell Class Reasoning Test, Form X

R.H. Ennis, W.L. Gardiner, R. Morrow, D. Paulus, & L. Ringel. (1964). Illinois Critical Thinking Project, University of Illinois, 1310 South Sixth Street, Champaign, IL 61820. (Grades 4-14)

The Cornell Conditional Reasoning Test, Form X

R.H. Ennis, W.L. Gardiner, J. Guzzetta, R. Morrow, D. Paulus, & L. Ringel. (1964). Illinois Critical Thinking Project, University of Illinois, 1310 South Sixth Street, Champaign, IL 61820. (Grades 4-14)

Cornell Critical Thinking Test, Level X

R.H. Ennis and J. Millman. (1985). Midwest Publications, P.O. Box 448, Pacific Grove, CA 93950. (Grades 4-14)

Cornell Critical Thinking Test, Level Z

R.H. Ennis and J. Millman. (1985). Midwest Publications, P.O. Box 448, Pacific Grove, CA 93950. (Advanced or gifted high school, college, adult)

The Ennis-Weir Critical Thinking Essay Test

R.H. Ennis and E. Weir. (1985). Midwest Publications, P.O. Box 448, Pacific Grove, CA 93950. (Grade 7-college)

Judgment: Deductive Logic and Assumption Recognition

E. Shaffer and J. Steiger. (1971). Instructional Objectives Exchange, P.O. Box 24095, Los Angeles, CA 90024. (Grades 7-12)

Logical Reasoning

A.F. Hertzka and J.P. Guilford. (1955). Sheridan Psychological Services, Inc., P.O. Box 6101, Orange, CA 92667. (High school, college, adult)

New Jersey Test of Reasoning Skills

V. Shipman. (1983). Institute for the Advancement of Philosophy for Children, Test Division, Montclair State College, Upper Montclair, NJ 08043. (Grades 4-college)

Ross Test of Higher Cognitive Processes

J.D. Ross and C.M. Ross. (1976). Academic Therapy Publications, 20 Commercial Boulevard, Novato, CA 94947. (Grades 4-6)

Test on Appraising Observations

S.P. Norris and R. King. (1983). Institute for Educational Research and Development, Memorial University of Newfoundland, St. John's, Newfoundland A1B 3X8. (Grades 7-14)

Test of Enquiry Skills

B.J. Fraser. (1979). Australian Council for Educational Research Limited, Frederick Street, Hawthorn, Victoria 3122, Australia. (Grades 7-10)

Watson-Glaser Critical Thinking Appraisal

G. Watson and E.M. Glaser. (1980). The Psychological Corporation, 555 Academic Court, San Antonio, TX 78204. (Grades 9-adult)

References on Critical Thinking Testing

Arter, J.A. and Salmon, J. (1987). Assessing Higher Order Thinking Skills: A Consumer's Guide. Northwest Regional Educational Laboratory, Evaluation and Assessment, 101 S.W. Main Street, Suite 500, Portland, OR 97204, pp. 15-31.

Ennis, R.H. and Norris, S.P. (In press). Critical Thinking Testing and Other Critical Thinking Evaluation: Status, Issues, Needs. In J. Algina (Ed.), Issues in Evaluation. New York: Ablex.

Norris, S.P. (1986). Evaluating Critical Thinking Ability. The History and Social Science Teacher, 21, 135-146.

Norris, S.P. and King, R. (1984). The Design of a Critical Thinking Test on Appraising Observations. St. John's, Newfoundland: Institute for Educational Research and Development, Memorial University of Newfoundland. (ERIC document Reproduction Service No. ED 260 083)

Walsh, D. and Paul, R.W. (1986). The Goal of Critical Thinking: From Educational Ideal to Educational Reality. American Federation of Teachers, 555 New Jersey Avenue N.W., Washington D.C. 20001, pp. 40-44.

Taken from: Test on Appraising Observations. Stephen P. Norris & Ruth King. St. John's, Newfoundland: Memorial University of Newfoundland, Institute for Educational Research and Development. (copyright, 1983)

PART A

A Traffic Accident

NORRIS HANDOUT

A traffic accident has just occurred at an intersection which has a stop sign in each direction. Several cars were involved.

A policeman and a policewoman will question people. Later several investigators will collect information about the accident. It is your job to judge the evidence given in the statements that follow.

1. A policeman is questioning Pierre and Martine. They were in their car at the intersection but were not involved in the accident. Martine is the driver and Pierre, who had been trying to figure out which way to go, is the map reader.

   The policeman asks Martine how many cars were at the intersection when the accident occurred. She answers, "**There were three cars.**"

   Pierre says, "No, **there were five cars.**"

2. A small boy and his father had been standing on the sidewalk when the accident occurred.

   The boy says, "**There was a motorcycle at the intersection.**"

   His father says, "No, **there was no motorcycle at the intersection.**"

3. A policewoman has been asking Mr. Wang and Ms. Vernon questions. She asks Mr. Wang, who was one of the people involved in the accident, whether he had used his signal.

   Mr. Wang answers, "Yes, **I did use my signal.**"

   Ms. Vernon had been driving a car which was not involved in the accident. She tells the officer, "**Mr. Wang did not use his signal.** But this didn't cause the accident."

4. The policewoman then points to Ms. Rosen's car which was one of the cars invol in the accident. She asks whether Ms. Rosen had signalled.

   Mr. Dawe, another driver not involved in the accident, says, "**Ms. Rosen signalled.** I was just talking to Ms. Vernon about this and I'm sure she will agree with what I said."

   Martine says, "**Ms. Rosen did not signal.** I'm sure I'm right."

5. The policeman talks to Mr. and Mrs. Peters, who were also involved in the accident. It is easy to see that Mr. Peters, who was the driver, is very upset by the accident. The policeman asks him to estimate his speed just before the accident.

   Mr. Peters says, "**I was going about 15 kilometers an hour.**"

   A little later when he is feeling well he says, "**I was going about 30 kilometers an hour.**"

6. The policeman asks whether or not the Peters' car had stopped at the stop sign. Ms. Vernon, who is a driver education instructor, says, "I am very experienced in these matters. **The Peters' car did not stop.**"

   Martine, who overheard this conversation, goes up to the officer and says, "**The Peters' car did stop at the stop sign.**"

7. The officer turns to question Martine and Pierre and Mr. Dawe. The officer asks them to estimate the speed of Mr. Wang's car when it hit the others.

   Mr. Dawe says, "**It was going about 40 or 45 kilometers an hour.**"

   The officer says, "It was going faster than that, wasn't it?" Martine says, "Oh yes, **it was going about 60 or 65 kilometers an hour.**"

8. Martine adds, "**Mr. Wang went right through the stop sign.**"

   The police officer turns to Mr. Dawe and says that at the scene of the accident Mr. Dawe couldn't remember whether Mr. Wang had stopped at the stop sign or not. Mr. Dawe says, "I remember now, **Mr. Wang did stop at the stop sign.**"

9. Ms. Vernon then says, "I also remember that **a fancy blue sports car went through the stop sign.**"

   Martine says, "**A car with twin headlights went right through the stop sign.**"

10. Martine says, "Three cars collided at the same time. **There was one crash.**"

    Ms. Vernon says, "**There was more than one crash.** It would be very strange for the three to collide at exactly the same time."

11. The police officers ask the people involved in the accident and the other drivers to come to the police station to make official statements. At the station, the policeman questions Mr. Peters.

    Mr. Peters points to a drawing of the intersection and says, "Just before the accident occurred **Mr. Wang's car approached the intersection from that direction.**"

    The police officer says to Mr. Peters, "Surely Mr. Wang's car came from a different direction." "Oh yes," says Mr. Peters, "**it did come from a different direction.**"

12. The policeman turns to Mr. Dawe to question him. In the background they can hear a conversation between the other officer and some of the other witnesses. Some are discussing whether one of the cars went through a stop sign.

    Mr. Dawe says, "**Mr. Wang and Ms. Rosen crashed into each other.** I saw it happen."

    "Also, I remember that **a car went straight through a stop sign**, too."

13. Nearby, the policewoman and Martine are looking at the drawing of the intersection.

    Martine says, "**A short time before the accident everyone was driving normally.**"

    She continues, "Then there was a loud squeal of tires. **Mr. Peters' car turned quickly toward the fruit stand on the corner.**"

14. The policewoman asks Mr. Dawe to tell in which direction Mr. Peters was travelling before the accident. Mr. Dawe says, "**He was going toward Fifth Street.**"

    The policewoman looks at her notes which were made at the scene of the accident. At that time Ms. Vernon had pointed and said that **Mr. Peters was going away from Fifth Street before the accident.**

In-classroom, Informal Assessment of Students' Thinking Skills

Kenneth Bumgarner

This session provided several examples of how classroom teachers could assess their students' thinking skills with non-paper and pencil, non-multiple-choice measurement methods. Ideologically, most of the methods presented were based on a model by Art Costa, which suggests the following 12 key behaviors be considered as evidence that student thinking is improving:

(1)	Persistence	time and effort put into task
		developing internal locus of control
(2)	Overcoming Impulsivity	planning ahead
		evaluating alternatives
(3)	Listening to Others	dialogical reasoning
		paraphrasing
(4)	Flexible Thinking	more accepting of ambiguity
		brainstorming
(5)	Metacognition	inner dialogue
		monitoring, adjusting thinking
(6)	Checking Accuracy	wait-time rather than immediate feedback
(7)	Questioning, Problem Posing	focusing on a goal
(8)	Drawing on Knowledge	assimilate, then apply
(9)	Precision of Language and Thought	trying for more accurate, thought out expression
(10)	Using all the Senses	e.g., draw watch face
(11)	Ingenuity, Originality Insightfulness	unexpected combinations
(12)	Wonderment, Curiosity	enjoyment of problem solving

Throughout this session the audience was actively involved in each of these 12 behaviors through a series of exercises -- illustrative of what a teacher could be practicing in order to stimulate thinking skills as well as of what behaviors should be looked for. Thinking was discussed as a process rather than simply requiring the students to arrive at an answer. A teacher should be more interested in observing behavior when the student does not know the answer, thereby being able to observe thought *processes*.

Several rather interesting examples were given throughout the session, in which audience participation was high. Analogy and strategy problems were presented, and techniques for responding to the students were demonstrated through use of the audience as a classroom. Some of the examples included several samples of "parlor games" activities, e.g.:

"It is true for butter but not for bread ..."

"A person may go through the door but not the window",

"I will gladly drink coffee but not tea"...

What is the relationship being expressed above?

Metacognitive thinking was discussed and illustrated and the importance of the affective domain to encourage the thinking process was emphasized. Using *all of the senses* in teaching thinking skills was illustrated several ways, e.g.:

"Try to draw (from memory) as close as possible the face of your watch." (I personally left the stem out of the picture and thought that I had four numbers which were actually only markers; my neighbor remembered quite a different set of characteristics.)

A "Thinking Log" was distributed at the end of the session, which can be used to stimulate immediate active processing to promote retention and the construction of knowledge as well as a record keeper for the sessions within a conference. Portions of the "Thinking Log" are attached.

THINKING ABOUT THINKING
CONFERENCE
SECOND THOUGHTS

SPONSOR:

Dr. Frank B. Brouillet, Superintendent of Public Instruction

CO-SPONSORS:

State Board of Education, Educational Service Districts, New Horizons for Learning, State Facilitator—National Diffusion Network, Association of Washington School Principals, Washington Association of School Administrators, Washington Association of Deans and Directors, Washington State Association for Supervision and Curriculum Development, Washington Education Association, Washington Roundtable, Washington State School Directors Association

Dear Participant:

This packet models a component of effective thinking--immediate active processing. Current research indicates that retention is enhanced and the understanding of concepts is clarified when such processing is employed as a learning strategy. The packet provides a structure within which you can reflect on and synthesize the information gained during the conference. The thinking log is a place for you and your team members to record and process your ideas, thoughts, questions and plans.

Please bring this log to each session. A block of time will be provided at the end of each session to allow you to analyze and record the essence of the presentation. Your notes will serve as a resource when you meet with your district team throughout the conference.

Additional "Reflection" Sheets have been provided in the back of this packet for the use of conference participants who are attending preconference and/or postconference workshops.

Jill Jacoby, Ph.D.
Director of Staff
Development/Curriculum
Educational Service District 113

Ken Bumgarner, Ed.D.
Director, Basic Education
Superintendent of Public
Instruction

Reflections

BUMGARNER
HANDOUT

1st General Session _____

A. List as many key words and phrases as you can recall about this presentation.

B. What three points do you consider most important.

1.

2.

3.

C. List three questions this presentation raised in your mind.

1.

2.

3.

BUMGARNER
HANDOUT

CONFERENCE PLANNER I

Schedule of Concurrent Sessions A
November 17, 1986

Presentation Title	Name of Team Member(s) covering the session

DISTRICT ANALYSIS: ALTERNATIVE ONE

If, like many districts, you are just beginning and have not yet started to work on thinking skills curriculum, you may wish to use this sheet to determine the current state of thinking skills in your school district.

1. How do you or does your district define thinking skills? Brainstorm and discuss suggested definitions.

2. Briefly describe how thinking skills are currently taught in your classroom, school or district. Indicate whether it has been infused across the curriculum.

3. What are some of the barriers to beginning or enhancing a thinking skills program in your area?

4. What are some strengths your district brings to the task of establishing a thinking skills program?

5. Other

DISTRICT ANALYSIS: ALTERNATIVE TWO

BUMGARNER
HANDOUT

If your district has worked on implementation of a thinking skills program over the last year, using your Reflection sheets as a resource, brainstorm the following questions to analyze your thinking skills focus:

1. Some important things we've learned over the past year in our school, classrooms, and district about teaching thinking.

2. The strengths of our present approach to incorporating thinking skills into classrooms and/or curriculum areas are:

3. Some additional components we should consider incorporating into our thinking skills program are:

4. In our district, training in thinking skills should focus on:

5. Other

DISTRICT PLANNING

November 18, 1986

Assuming that your district is interested in pursuing continuing a thinking skills program, briefly summarize your committee's recommendations as to how your district might proceed.

Step 1.

Step 2.

Step 3.

Step 4.

Step 5.

Step 6.

Use the next sheet to help you map your plans.

What materials/resources will you need?

Estimate the cost of your proposal.

Materials

Consultant Fees

Teacher Inservice

Teacher Release

Total

BUMGARNER
HANDOUT

PLANNING TEAM TASKS AND RESPONSIBILITIES

	CALENDAR		OBJECTIVES	PERSON/S RESPONSIBLE
	Start	Finish	Describe tasks	Name and phone number of persons responsible for each task
1.				
2.				
3.				
4.				
5.				
6.				

Sow's Ears Into Silk Purses -- How to Take Unlikely-Looking Textbook Material and Make It Testable for Critical Thinking

Connie Missimer

Ms. Missimer's presentation was based on the premise that we should use the structure of "theory" as our framework (theory) out of which to teach and assess critical thinking. Critical thinking is the making and appreciating of good theories or arguments. Arguments, theories, hypotheses, problems with their solutions, even sales pitches--all have the same basic structure or framework. Ms. Missimer believes that it is the use of alternative theories which drives the intellectual life and causes progress in all subjects.

Missimer presented a visual way to represent theories, using the simple drawing of a house. The ground is the problem, the walls are the reasons which support the roof, which is the conclusion, assumptions are below ground in the basement. Just as there are many houses on a street, so there are many possible solutions to a problem, but each must be supported with reasons. By helping students to see the structure of their arguments, and allowing a person to change arguments (build a new structure), we can incorporate critical thinking vocabulary, skills, and dispositions into all parts of the curriculum where teachers invoke this structure.

Missimer showed how a play such as "Romeo & Juliet" can be much more interesting for students if they are offered to theorize about whether the play is any good, rather than to memorize ways that it is good. In the latter case, students are memorizing one theory at best, and are not considering alternative hypotheses. The latter is conducive to independent, critical thinking, and besides is more enjoyable.

American Government textbooks are a challenge, since they are often made up of separate bits of information. Missimer indicated that even these texts can be used by teachers and students to create theories or arguments. For example, high school seniors could grapple with the question "What political institutions are **missing** from the table of contents?" (Hint: What political institutions are present in other countries, but don't appear in ours?") As students see that our political institutions don't include religious leaders, the military, etc., they come to understand the significance of the American system much better in contrast to other systems (which are, of course, theories). It is this "steroscopic" view of political systems, or indeed of any subject, which alternative theories provide.

Missimer also showed an example of a test of higher order thinking which she designed. It is being used in a few high schools and colleges, although no technical data about its efficacy are available. The first section of the "Test of Critical Thinking Skills" offers arguments of increasing complexity. The multiple-choice questions ask "Which sentence is the conclusion?" "What reason(s) is offered?" "What is the issue?" The attempt here is to ascertain how well students understand an argument when it is presented to them. The second part consists of unfinished dialogues by two opposing parties. The student is asked which statement best continues the dialogue. (Statements other than the correct one are either contradictory or irrelevant.) The third section of the test offers a description of four major types of evidence: experimental, correlational, speculative, and observational. Students are asked to tell which type of evidence they should most expect in four different subject areas.

The "Test of Critical Thinking Skills" is meant to improve on current tests (Cornell Critical Thinking Test and Watson-Glaser Critical Thinking Appraisal) by avoiding having students make inferences that are not directly spelled out in the

passages. It also does not require prior academic knowledge. Like Cornell and Watson-Glaser, this test does not assess how creative a critical thinker the student is.

Missimer's test and handouts are attached.

Test of Critical Thinking Skills

PART I

You will be asked to identify the issue, conclusion, and reason(s) in the following arguments. The conclusion is the main, overall point being made. The reason or reasons are the ideas in support of the conclusion. The issue is the question about which the argument takes a stand. **Please remember that in an argument the conclusion is not necessarily the last sentence.**

Driving while low on gas is a bad idea. First, dirt particles at the bottom of the gas tank will be pushed into the motor with the remaining gas. Second, there is a danger of running out of gas, either because the gas guage is inaccurate or because the driver forgets to fill up the tank.

1. The conclusion of the argument is:
   a) dirt particles at the bottom of the gas tank will be pushed into the motor
   b) there is a danger of running out of gas
   c) driving while low on gas is a bad idea
   d) the driver forgets to fill up the tank

2. A reason offered to support the conclusion is that
   a) there is a danger of running out of gas
   b) many drivers are careless
   c) driving while low on gas is a bad idea
   d) gas costs less now than it did in the recent past

3. The issue being argued is
   a) whether driving while low on gas is a bad idea and if so, why
   b) whether drivers are careless and if so, why
   c) whether gas costs less than it did in the recent past and if so, why
   d) whether dirt particles get pushed into the motor and if so, why

Stealing is wrong. But the most important reason that it is wrong is not what one might think, for instance that the person stolen from has been unjustly deprived of some things. Rather, stealing is wrong mainly because of the injury that it does to the character of the thief.

4. The conclusion of this argument is:
   a) that the person stolen from has been unjustly deprived of some things
   b) that stealing injures the character of the thief
   c) that stealing does harm
   d) that stealing is wrong

5. The reason supporting the conclusion is:
   a) that the person stolen from has been unjustly deprived of some things
   b) that stealing injures the character of the thief
   c) that stealing does harm
   d) that stealing is wrong

6. The issue being argued is
   a) whether the person stolen from has been unjustly deprived of some things and if so, why
   b) whether stealing injures the character of the thief and if so, why
   c) whether stealing does harm and if so, why
   d) whether stealing is wrong and if so, why

"Large-scale organization does not always diminish the individual's freedom– in some respects it enlarges it.

As a result of large-scale organization modern man enjoys freedoms that he could not hope to enjoy otherwise. Out of the vast and elaborately organized medical center come findings that free the individual from illnesses that have plagued mankind for centuries. The great urban university, which strikes some critics as little more than a huge factory, places within reach of millions of low-income workers the opportunity to surmount ignorance and stretch their horizons.

The man who moves from a small town to a large city experiences unaccustomed freedom. He not only escapes the stultifying web of attitudes, expectations and censorship that characterize the small town, he finds in the city more choices in every dimension— kinds of dwelling, consumer goods, entertainment, social companions, culture and work."

John Gardner, *Self-Renewal*, p.76

7. Which of the following conclusions would go directly against Gardner's point of view?

a) Large-scale organizations invariably diminish the individual's freedom.

b) There should be a larger pool of consumer goods from which to choose

c) Culture and workplace are more interrelated than has previously been suspected.

d) Fewer illnesses plague mankind now than in past centuries.

8. Which of the following are major reasons which Gardner offers for his conclusion?

a) big urban universities are little more than huge factories; many people have moved to the city

b) vast medical centers free individuals from illness; big urban universities offer the poor a chance to learn

c) people in small towns are unaccustomed to the freedom of big cities; society needs more kinds of dwelling, consumer goods, entertainment, social companions, culture and work.

Here's the dilemma: Is it better to go to the market once a week or more often? Going once a week saves time and gas; on the other hand, going more often enables one to have fresher vegetables and to take advantage of more sales.

9. The issue is
   a) whether it is better to go to the market once a week or more often
   b) whether going once a week saves time and gas
   c) whether going more often enables one to take advantage of more sales
   d) the issue is not made clear

10. The conclusion is
   a) it is better to go to the market at least once a week
   b) it is better to go to the market only once a week
   c) it is better to go to the market more than once a week
   d) the conclusion is not made clear

11. The reasons are
   a) all in favor of going to the market only once a week
   b) all in favor of going to the market more than once a week
   c) two reasons favor going once a week and two reasons favor going more often.
   d) it is not clear what the reasons are

The group, "First Amendment Rights for Children," makes the argument that schools should not censor books for children. A child should be able to go to the school library and request any book he wants, since children are citizens and therefore should enjoy all reasonable constitutional rights.

One counterargument has been raised that school libraries should censor books for children. Some books should not be made available to children because they could be harmed by not understanding the full context of these books. Take Truman Capote's *In Cold Blood* , for instance. This is an account of two men who murder a family. Capote is trying to understand the killers, but a child might believe that he is advocating murder. Therefore children ought not to be allowed to read a book like this until they are in high school.

This counterargument is, however, incomplete. It does not take two important matters into consideration. First, grade schools have a limited amount of time to instill basic skills and ideas into students. Therefore, grade school administrators should not order any books that in their judgment are *a waste of time* , i.e., do not help to instill the basic skills and ideas that children need to learn. Second, schools don't have unlimited funds for school libraries, and so should order only those books which are instructive and enjoyable to children. Schools, therefore, should censor any books that in their judgment do not conduce to the best expenditure of time and money.

12. The conclusion is that
    a) there are many arguments involved in the question whether grade school libraries should censor children's books
    b) censorship of children's books by school libraries is justified
    c) children's books should not be censored, because the First Amendment guarantees them the right to read what they wish
    d) grade schools are not doing a good enough job censoring students' reading material.

13. The reasons are that
    a) children could be harmed by not understanding the context of some books.

b) the child only has a limited amount of time to absorb basic skills and ideas
c) the school has only a limited amount of money for library books
d) all of the above
e) none of the above

14. The argument assumes
    a) that children need not have the same rights as adults
    b) that children have the same right to informed consent as adults
    c) that school authorities are unfair

END PART I

PART II Continuing Dialogues

The following is an unfinished dialogue between two people. You will be asked to anticipate a further remark that would be consistent with the person's remarks so far.

Hal: There's nothing to do here on campus except go to class.

Mollie: I'll have to take issue with you, Hal. This week alone there was a jazz concert, a gallery show and three lectures by visiting professors.

Hal: I see that I should have drawn a distinction between those activities you rightly point out are extracurricular, and those things which I enjoy; I've never seen the latter on campus.

Mollie: Granted that you've never seen notices for extracurricular activities that you know you enjoy--- is it possible that you are falsely assuming that you would not enjoy the activities we now have? How do you know you won't enjoy some of these school activities unless you try them?

15. Which of the following could Hal say to further his line of reasoning and remain consistent with what he has said so far?

Hal: a) Well, there are many extracurricular activities on campus that interest me. I just haven't had the time to attend them.

Hal: b) I don't believe that I am making a false distinction between activities I consider fun and those which the school supports. In fact, friends who have attended some of these school activities have told me about them, and I'm fairly sure they would bore me.

Hal: c) I am operating on the assumption that if you set your mind to it, you can enjoy anything. And such is the case with school activities, whether they be of an extracurricular or academic nature.

Likewise, follow this new dialogue carefully so that you can add to Anne's line of reasoning.

Anne: Students should have to take basic courses in all major subject areas.

Brad: I disagree. Forcing students to take certain courses destroys their natural incentive to learn.

Anne: But it's claimed that most people will change jobs seven times during their lifetime. If colleges don't prepare students for a diversity of occupations, then they are failing in their duty to students.

Brad: You are assuming that four years of college are supposed to prepare a student for all contingencies. I think that the purpose of college is to get students excited about learning in the hope that they will return to college throughout their lives for more education.

16. Which of the following would logically extend the dialogue?

Anne: a) You are assuming that making students take subjects will prevent learning excitement

Anne: b) You are assuming that students should take some of the same subjects before graduation.

Anne: c). You are assuming that if students have to take certain courses they will become enthusiastic about subject areas which they would otherwise probably never know about.

END PART II

PART III Evidence

We get evidence to support conclusions through these four means: scientific experiment, correlation, speculation, or direct observation.

1. **Scientific experiment** -
   Scientific method involves setting up an experiment as well as the control (an almost identical situation, except that the variable to be tested in the experiment is missing.)
   - Scientific method produces the strongest evidence, but it works only if all variables can be controlled.
   When that is not possible, the next best evidence is correlation.
2. **Correlation** -
   With a correlation one can see that there is a relationship between two variables. The evidence that one variable is the reason for the other (conclusion) is not as strong as in scientific method. But since we cannot hold down all variables in many situations, a correlation is the best evidence we have.
3. **Speculation** -
   When we cannot set up a scientific experiment nor find any correlations yet still believe a theory to be true, we must justify our belief by an appeal to others' reason and experience. This is the area of speculation.
4. **A single observation** -
   Sometimes the best evidence we can obtain is an eye-witness account.

Any of these types evidence could be obtained badly and thereby be faulty or fraudulent. But assume here that whatever the type of evidence which could be obtained, it would be of the highest quality. With that presupposition in mind, which of these types of evidence would be the strongest that you could expect in the following:

Example:
A theory that when people exercise they decrease their chance of a heart attack.
a. experiment b. correlation c. speculation d. a single observation

The correct answer is b. It would not be ethical to hold people in cages and force some to exercise, others to rest. So experiment is out. The best we can do is to correlate the incidence of heart attack to the amount of exercise in a large number of people.

Choose among the four types of evidence—experiment, correlation, speculation, a single observation—and determine which of these would be the strongest that you could expect in the following circumstances:

17. A theory that war is never justified.
    a. experiment
    b. correlation
    c. speculation
    d. a single observation

18. A theory that plants do not grow in response to sound.
    a. experiment
    b. correlation
    c. speculation
    d. a single observation

19. A theory that when the dollar drops in value, the stock market rises.
    a. experiment
    b. correlation
    c. speculation
    d. a single observation

20. A court case that a certain individual committed a murder.
    a. experiment
    b. correlation
    c. speculation
    d. a single observation

END PART III

Answers:
Part I 1c, 2a, 3a, 4d, 5b, 6d, 7a, 8b, 9a,10d, 11c, 12b, 13d 14a

Part II 15b, 16a,

Part III 17c, 18a,19b, 20d

Test prepared by C.A. Missimer, c 1987

MISSIMER
HANDOUT

Pedagogical Schemes for Critical Thinking:

How do you pull it out of texts? How can you test it?

For the conference, "Assessing Higher Order Thinking Skills," Oct. 2, 1987

By Connie Missimer

First briefly, what is critical thinking?

I would claim that **critical thinking is the making and appreciating of good arguments.**

Synonyms for arguing: hypothesizing, theorizing, making the case that...

Unlike narratives, which proceed by a temporal thread, arguments are structures (see diagram, last page)

What specific kinds of questions will produce critical thinking ?

Questions that hold up competing theories or arguments.

1. **Argue that theory X (X) is preferable** to theory Y (Y).

e.g.,

Argue that equality of opportunity is preferable to equality of results.

Argue that Darwin's theory of evolution is preferable to Lamarck's.

Argue that living in 17th century France is preferable to living in the U.S. today.

Tell me why the boy in the story should have stayed with the Pooh Bear instead of going right home. Now tell me why the boy should have gone right home. Now, what do you think he should have done? How come? (fun-fun versus fun-serious reasons

Your example(s)

_____

_____

or

Further the argument that the text is making that X is preferable to Y; **or oppose** that argument.

_____

_____

2. Argue for and against_____

e.g., Argue for and against the idea that Japanese competition is a decisive factor in the slowing of U.S. economic growth in the '80's.
Give me reasons why we should or should not have a spelling test this week.

_____

_____

3. **Create a theory that would explain why** _____
   e.g., Create a theory that would explain why Milton's most attractive character in "Paradise Lost" is Lucifer.
   or Create an idea of why the girl could have lost her hat. (An reason John gives becomes "John's theory" in further discussion.)
   You want students to hunt for a variety of reasons.

4. What conclusions could you draw about (e.g.,Hardy's view of life, stellar motions, the economic outlook) from these
   features in Hardy's novels: harsh landscapes, lonely people, etc.
   phenomena recently sighted at Palomar:
   recent **events** in the U.S. economy:
   Or If the author of the book says that _____, what do you think he means (or is concluding)?
   Since I found papers on the hall floor, a banana peel on the front stairs, and spilled juice in the bathroom, what could I conclude about your tidiness? (And you hope for a range of answers so you get to say) And which conclusion/inference is more justified, in your opinion?

The features, phenomena, or events the book or you cite are the reasons for your students' and your theoretical conclusions. You want them to play around with various inferences.

_____

_____

4. (harder) What **significant fact (value) is missing** in _____?
e.g., What value was missing from popular culture in the U.S. in the 1970's? Make an argument about the implications of the lack of this value. Or Create an argument that this value is (is not) central to a healthy society.

Or What could we in the United State have considered before we started using a lot of electricity?
When Marian got mad and hit Joey, what other way of behaving do you think might have been missing from her mind?

Questions that are not immediately conducive to critical thinking (but which may be quite important for other purposes!)

Can you summarize, list, name the.....
Just "compare and contrast" do not do the full job, because they do not offer reasons, just conclusions.

Questions which might be conducive to CT, but which seem ambiguous in that regard:

What four x's are involved in process y?
What is the significance of.....
Discuss..... and its relevance to.....

Connie Missimer
4836 N.E. 40th St.
Seattle, WA 98105
(206) 522-4985

BEST COPY AVAILABLE

Leaps are necessary but some leaps are safer than others. Usually used to describe an unsafe one: "In suggesting that...on the basis of just a few cases, the author has made too great an inductive leap."

Infer Come to the conclusion that. "From the look of his sweaty body one can infer that he has been doing strenuous exercise," or "We can make the inference that."

Inference The conclusion. "Although he thinks it's right to conclude that we should be patriotic, he doesn't think it right to base this inference on the premise that we are superior to people of other nations."

Issue The topic or subject at hand.

Make the argument Same as "one could argue that." "One could make the argument that."

Necessary cause A factor without which some effect would not occur. "Oxygen is a necessary cause of fire" (no oxygen, no fire). It often takes several necessary causes to create an effect (e.g., to get a rocket into space, or to get to work on time). *See also* Sufficient cause.

Necessarily Must be the case (e.g., "necessarily involves").

Not necessarily Doesn't have to be the case. "Because some people are better at argumentation doesn't necessarily imply (indicate) that they are smarter. These better arguers have probably just been practicing longer"; often found in the phrases, "does not necessarily imply," "is not necessarily the case."

Ought Usually signals the solution a thinker has arrived at as a result of his conclusion. "Because of the rapid rate at which we're polluting this lake, I conclude that we will not be able to swim in it by 1996. So we ought to find a new lake to pollute (just kidding)." *See also* Should.

Play the devil's advocate Take the opposite point of view. Student to instructor: "I'd like to play the devil's advocate to your position that our class should have an in-class final. Think of all the extra love and attention we could give to your questions if we had lots of time to mull over them at home."

Premise Reason. *See also* Based on the premise that.

That presupposes That requires the assumption (underlying reason) that. "That argument presupposes that everyone is born with equal enthusiasm and talent."

Reason The support for the conclusion; the why. Write, "The reason is," don't write, "The reason is because," because that's redundant. "The reason for her tardiness is her laziness," or "The reason that she is often tardy is *that* the bus schedule is erratic."

Should Signals a solution or a prescription (as opposed to a statement about what the case is). *See also* Ought

Sufficient cause A cause that by itself will create an effect. "Oxygen is a necessary cause of a fire, but it isn't sufficient. On the other hand, lack of oxygen is a sufficient cause of choking to death."

Therefore Signals that a conclusion must follow and that a reason came before (usually right before, but not always). "She's always up at 4 A.M. Therefore, she's dead tired by 10 P.M."

Warranted inference A phrase to the effect that a conclusion and/or the step from the reason(s) to the conclusion is justified. "Many studies have found a positive correlation between the amount of time students spend on a course and their grades. These studies warrant the inference that sheer effort expended is a central factor in doing well in school."

MISSIMER HANDOUT

...e following is a list of frequently used phrases in argumentation, ...h some illustrative examples of their use.

...ree to disagree An accord that arguers are at an impasse and shouldn't keep arguing, probably because they will argue in a circle. Arguers should agree to disagree when (1) they need to get more factual evidence, if the argument hinged on a fact, or (2) they need to come up with different approaches or expanded arguments.

...sumption The reason beneath the stated reason(s). Either the author will not state them at all, or the author will announce them by saying, "I am assuming that," or "The assumption here is."

...sed on the premise that "Based on the reason (or assumption) that."

...sed on the presupposition that Same as "based on the premise that," above.

...cause Indicates that a conclusion has preceded and a reason will follow.

...se See chapter 12.

...im Any statement. Claims are provisional, subject to change with new information.

...duce To arrive at a particular instance that is entailed in a generalization. Often found in the form, "from which we may deduce that." "All our parachutes open when the rip cord is pulled, from which you may deduce that the one you buy from us will open too."

...finition All that is meant to be included in a term.

...tinction (Between two terms) An important difference between two terms.

...w the inference that Come to the conclusion that.

...w the conclusion that Same as "draw the inference that."

...n granting that...still In a counterargument, a technique for accepting or granting the truth of the other person's reason(s), and yet showing how they are consistent with your argument. "Even granting that it is frightening to speak out in a group and that people may think you're stupid, still it is better to speak your mind because the world needs your ideas just as much as it needs anybody else's."

...se dichotomy between In a counterargument, the claim that a false distinction has been made. He: "In college, you can either have a lot of fun or you can do a lot of studying." She: "You have created a false dichotomy between having fun and studying. By George, you can have a lot of fun studying!"

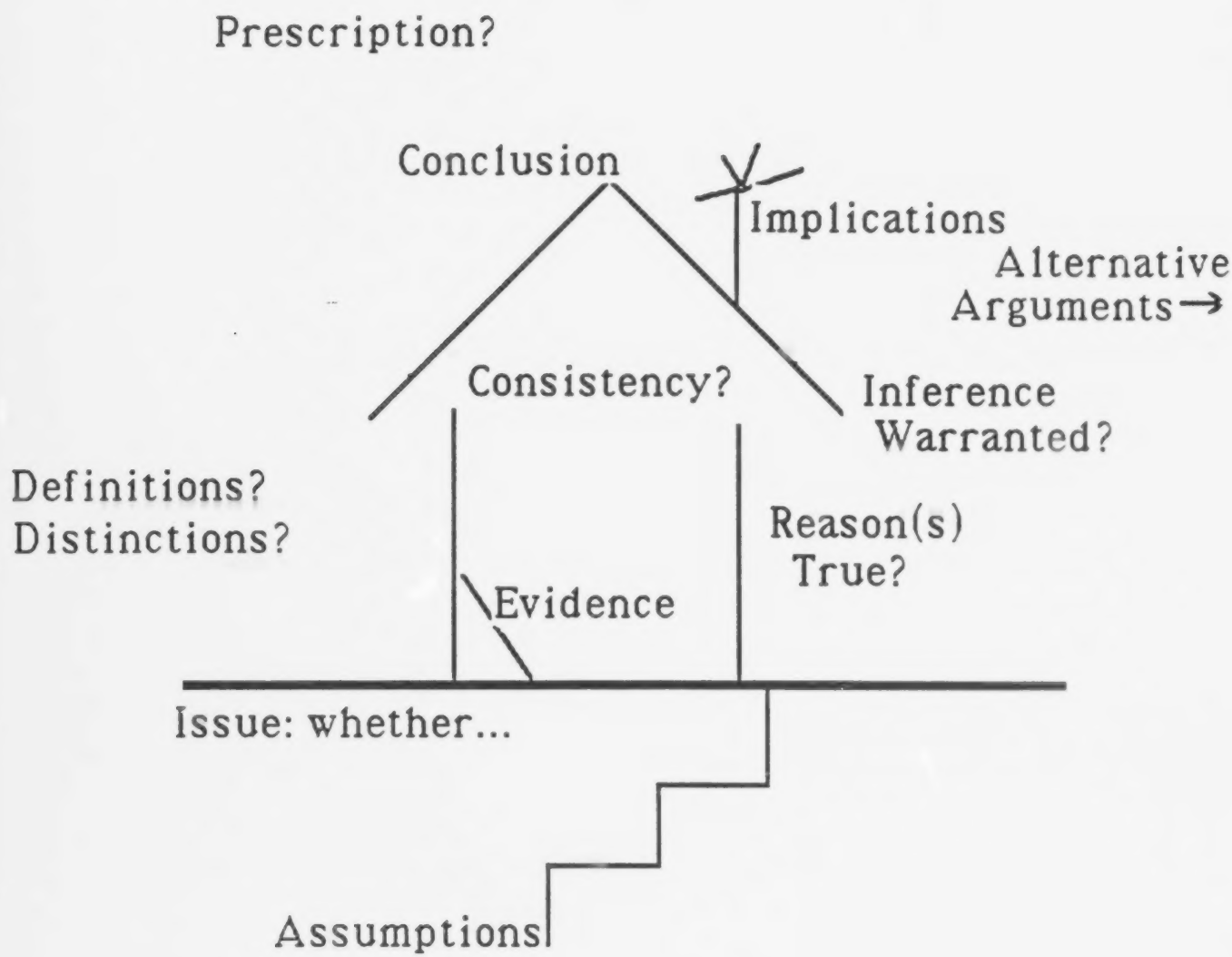
...ages on Usually the argument hinges on, or is dependent on, some outside fact or other argument (e.g., "The argument that most societies have suppressed women hinges on the claim that there is no biological explanation for women's failure to hold positions of power").

...then What follows "if" is the reason; what follows "then" is the conclusion to that statement.

...plications Consequences beyond the conclusion. Either the author does not state them or the author says, "The implications are."

...ply Most frequently found in a counterargument or used by a respondent in a deliberation. It is a request to clarify someone else's conclusion. "You seem to be suggesting a connection between the cultural values held by a group and their average income. Are you implying that most wealth and poverty are the result of these values?" Or you can use it to say what you don't mean: "But I don't mean to imply that." A less frequent but handy use: One idea necessarily involves another (e.g., "Drama implies conflict").

...ction The inference to a general conclusion from a set of particulars



The Structure of Argument

C.A. Missimer, *Good Arguments*
(c) Prentice-Hall, 1986

How To Select A Test Of Higher Order Thinking Skills

Judith A. Arter

The goal of this session was to promote good consumerism by discussing the features that tests of HOTS should have, and comparing currently published assessment devices to this ideal list. The purpose was not to criticize any particular tests or imply that we should not use multiple-choice tests; but, rather, to describe assessment issues and concerns so that users know the strengths and weaknesses of what they are using. The list of features was taken from two appendices **Assessing Higher Order Thinking Skills: A Consumer's Guide** published recently by NWREL. These appendices were reproduced as a handout for this session and are attached.

Basically, the checklist for selecting tests of HOTS reflects the issues and concerns that were raised during the keynote address and responses by the panel. These issues, and how they relate to current tests are:

A. **Content**. The first concern is to select a test that looks like it measures what you want it to measure. Not all tests of HOTS attempt to measure the same thing.

1. Tests can have vastly different emphases. The first four columns of the attached chart describe the content of the tests. There were tests found that covered critical thinking, problem solving, development (e.g., Piaget), creativity, structure of the intellect and Bloom's Taxonomy (e.g., achievement tests).

2. Even when tests are called the same thing, what they cover can be different (e.g., Watson-Glaser v. Cornell Critical Thinking Tests).

3. Does the test measure HOTS in an atomistic or holistic manner? Most tests are based on taxonomies of HOTS where each question is designed to measure one skill. There are few that look at how all the skills work in concert.

4. Are the questions abstract or is the attempt made to have them reflect real-life situations? Some tests (e.g., the structure of the intellect tests) can have questions which are very abstract -- e.g., analogies, number sequences, syllogisms, and pattern recognition. The rationale for the item types needs to be clear, and needs to be what you want to emphasize.

   One problem is that the more "real" a situation becomes, the less easy it is to have questions with only one right answer. Abstract items generally have only one clearly right answer.

5. Is the test content embedded in a subject area or is it designed to be "subject free?" Examples were provided of situations in which the claim of the test to be embedded in a subject area were not entirely accurate.

6. Is the test based on some theoretical model? Most of the tests are. However, as was pointed out during the opening session, not all theoretical models are equally good.

B. **Usefulness.**

1. Is the instrument or method easy to use? Most tests are group, multiple-choice tests. These are generally easy to use. Some measures are individually administered or open-ended. These tests are generally much harder to give and score. So, content advantages that might arise from having more of an open-ended test are balanced by having them harder to use.

   Column 8 of the chart provides information about test format; columns 9 and 10 give information on how long it takes to administer the test.

2. Is there help with interpreting the results and using the results for instructional planning? Columns 11 and 12 of the chart provide information on these points. In general, the tests are lacking on this dimension. There are some that are tied directly into curriculum materials.

C. **Reliability.** Test reliability refers to the consistency with which a test measures what it does. Column 13 provides information about the reliability of various tests. In general, the total-test score reliability of most tests is reasonable. However, subtest score reliabilities tend to be very low.

D. **Validity.** Validity is a key issue. What evidence is there that the test measures what it claims to measure? The test should provide evidence that scores are due to HOTS and not due to extraneous factors. Below are listed some of the studies that ideally would be done on a test. Column 14 of the chart discusses validity.

1. There should be a effort at content validation. Part of this is reviews of the literature and relationship to a theoretical model. Another aspect might be review of questions by knowledgeable judges -- do judges agree that the questions measure the skills claimed? Most tests have adequate review of questions.

2. Is the test measuring HOTS or vocabulary? The students need to understand what is being asked or the test does not measure HOTS. Most tests do not see if this is the case.

3. Right answers are arrived at only through the thinking process being claimed and not for other reasons such as differences in general knowledge, cultural differences or content area knowledge. The presenter illustrated these points with examples from several tests. Most tests do not examine this.

4. Correlations are presented between test scores and scores on other measures, and there is some rationale for why the correlations presented support the validity of the test. Many tests report correlations but do not explain why they support its validity.

5. What is the evidence to support the subtest structure of the tests. Many of the tests claim to measure separate subskills. Are they really different? A factor analysis is sometimes done to examine this issue. The structure of the intellect tests are generally best at this because that is where the structure comes from.

6. Is there other evidence of validity such as logical group differences in scores or differences in scores after instruction? These are also not usually done by tests.

7. Does the test discuss these measurement issues and how it has attempted to deal with them? Most tests do not even acknowledge that they exist. Two tests that are good in this regard are the Test of Appraising Observation and the Cornell Critical Thinking Test.

In general, there are many assessment issues regarding HOTS. Some of these are equally relevant to instruction -- definitions, reductionism, and the degree to which they are embedded in content areas. Current commercially available assessment devices are of variable quality. Test developers are, however, continuously attempting to attack these issues as new tests are developed.

Northwest Regional Educational Laboratory
101 SW Main Street, Suite 500
Portland, Oregon 97204
Telephone (503) 275-9500

HIGHER ORDER THINKING SKILLS TESTS

Test Center
800/547-6339
503/275-9570

From: Assessing Higher Order Thinking Skills: A Consumer's Guide by Judith A. Arter and Jennifer Salmon, Portland, OR: Northwest Regional Educational Laboratory, 1987.

Summary Table of Instrument Characteristics

Instrument	Focus	Grades	Subject Specificity	No. Forms	No. Levels	No. Items	Item Type	Adm. Time	Scoring	Norms	Other Interp.	Reliability	Validity	Comments	Availability
CRITICAL THINKING AND PROBLEM SOLVING:															
Applications of Generalizations Test (1969)	Generalizations	4-12	Social Studies	1	1	65	M.C.	?	Hand	Fair	None	Fair	Some+	Specifically tied to Taba curriculum	Tests in Microfiche (#020842611), ETS Test Collection, Princeton, NJ 08541-0001
Cornell Class Reasoning Test (1964)	Class Reasoning	4-12	General	1	1	73	M.C.	40 min	Hand	Fair	None	Fair	Some	Structured formal logic only	Illinois Critical Thinking Project, Univ of Illinois Champaign, IL 61820 or ERIC ED 037818
Cornell Conditional Reasoning Test (1964)	Conditional Reasoning	4-12	General	1	1	72	M.C.	40 min	Hand	Fair	None	Fair	Some	Structured formal logic only	Illinois Critical Thinking Project, Univ of Illinois Champaign, IL 61820 or ERIC ED 037818
Cornell Critical Thinking Tests	Critical Thinking	X 4-12 Z adult	General	1		71 52	M.C.	60 min	Hand, Machine	Fair	Some	Fair-Good	Some		Midwest Publications PO Box 448 Pacific Grove, CA 93950
Ennis-Weir Critical Thinking Essay Test (1985)	Critical Thinking	9-adult	General	1	1	1	Essay	70 min.	Hand	Fair	Some	Fair-Good	Some	Scoring requires training.	Midwest Publications PO Box 448 Pacific Grove, CA 93950
Judgment: Deductive Logic & Assumption Recognition (1971)	Critical Thinking	7-12	General	1	1	135	M.C.	135 min	Hand	None	None	No Info.	None	Has five separate aspect specific tests.	IOX Assessment Associates Box 24095 Los Angeles, CA 90024
Means-Ends Problem Solving (1975)	Interpersonal Problem Solving	5-7 9-adult	General	1	2	10	Essay, Interview	None given	Hand	Fair	Some	Good-Excellent	Extensive	Scoring and administration requires training.	Department of Mental Health Science, Hahnemann Univ., 112 N. Broad St. Philadelphia, PA
New Jersey Test of Reasoning Skills (1985)	General	4-adult	General	1	1	90	M.C.	70-80 min	Machine only	Fair	None	Good-Excellent	Some	No reliabilities are reported for sub-skills. Test is rented from publisher. Tied to the Philosophy for Children Program	Institute for the Advancement of Philosophy for Children, Montclair State College, Upper Montclair, NJ 07043
Primary Test of Higher Processes Thinking (1978)	Bloom's Taxonomy	2-4	General	1	1	55	M.C., open, match	60 min	Hand	Fair	None	Fair	None		Tests in Microfiche (#011161), ETS Test Collection, Princeton, NJ 08541-0001

ARTER HANDOUT

Instrument	Focus	Grades	Subject Specificity	No. Forms	No Levels	No Items	Item Type	Adm Time	Scoring	Norms	Other Interp	Reliability	Validity	Comments	Availability
CRITICAL THINKING AND PROBLEM SOLVING (continued...)															
Purdue Elementary Problem Solving Inventory	Problem Solving	2-6	General	1	1	49	M.C.	40-50 min	Hand	Fair	None	Fair	Some	Uses a filmstrip and tape to present questions	Illinois Educational Research Institute, Purdue Univ., S. Campus Courts, Bldg. D, W. Lafayette, IN 47907
Ross Test of Higher Cognitive Processes (1976)	Bloom's Taxonomy	4-6	General	1	1	159	M.C.	140 min	Hand, Machine	Good	Some	Excellent	Some	Scoring one by one consuming	Academic Therapy Publications, 20 Commercial Blvd. Novato, CA 94947-6191
TAB Test. An Inventory of Science Methods (1966)	Problem Solving	4-6	Science	2	1	2	Performance situations	?	?	Fair	Some	Poor	Some	Student chooses sequence of 'experiments' to answer a question by pulling tabs to uncover results	Tests in Microfiche (#007741), ETS Test Collection, Princeton, NJ 08541-001
Test of Enquiry Skills (1979)	Critical Thinking	7-10	Science	1	1	87	M.C.	1/2 - 1 hours depending on grade	Hand	Fair	Some	Fair	Some	Only one subtest pertains to critical thinking.	Australian Council for Educational Research, Frederick St., Hawthorne, Victoria 3122, Australia
Test of Science Comprehension (1963)	Interpreting Data	4-6	Science	1	1	50	M.C.	90 min	Hand	Fair	Some	Fair	Some		A Test of Science Comprehension for Upper Elementary Grades, Science Ed. 47, 319-330 (1963).
Test on Appraising Observations (1983)	Appraising Observations	7-adult	General	1	1	50	M.C.	50 min	Hand	Fair	Some	Poor	Extensive		Institute for Educational Research and Development, Memorial University of Newfoundland, St. Johns, Newfoundland, Canada A1B3X8
Think It Through (1976)	Problem Solving	preK-1	General	1	2	51	M.C.	30-40 min	Hand,	Good	Extensive	Fair	Some	Questions are read to students	CTB/McGraw Hill Del Monte Research Park, 2500 Garden Road, Monterey, CA 93940
Watson-Glaser Critical Thinking Appraisal (1980)	Critical Thinking	9-adult	General	2	1	80	M.C.	40 min	Hand, Machine	Good	Some	Fair Good	Extensive		Psychological Corporation, 555 Academic Court, San Antonio, TX 78204-0952

ARTER HANDOUT

Instrument	Focus	Grades	Subject Specificity	No. Forms	No. Levels	No. Items	Item Type	Adm. Time	Scoring	Norms	Other Interp	Reliability	Validity	Comments	Availability
DEVELOPMENTAL TESTS:															
Arlin Test of Formal Reasoning (1984)	Piaget	6-adult	General	1	1	32	M.C.	45 min	Hand, Machine	Fair	Some	Poor	Some	Follow up questions ask for reasoning behind answer. Hand scoring is awkward.	Slosson Education Publications, P.O. Box 280, E. Aurora, NY 14052
Classroom Test of Formal Reasoning (1978)	Piaget	8-12	General	1	1	15	M.C.	75-100 min	Hand	Fair	Some+	Fair	Some+		The Development and Validation of a Classroom Test of Formal Reasoning, Journal of Res in Science Teaching, 5, 11-24 (1978)
Formal Operations Measure (no date)	Piaget	Adult	Science	2	1	7	Open-ended	45-60 min	Hand	Most	Some	Most	Most		Tests in Microfiche (#09817?), ETS Test Collection, Princeton, NJ 08541-0001
Formal Operations Test	Piaget	8-adult	Biology, History, Literature	1 each	1 each	30 each	M.C.	?	Hand	Fair	Most	Most	Some	Three separate tests.	ETS Test Collection, Tests in Microfiche (8422, 8423, 8424), Princeton, NJ 08541-0001
Springe Task (1978)	Piaget	5-adult	General	1	1	19	Indiv. Open-ended	15 min.	Hand	?	?	Good	Some	Requires an apparatus	A Measure of Scientific Reasoning: The Springe Task, J. Ed. Measurement, 15, 1978, ERIC ED 169971
Test of Logical Thinking (1979)	Piaget	6-adult	General	2	1	10	M.C.	30 min	Hand	Fair	Most	Good	Some+	Examinee pick answer and justification	Kenneth Tobin and William Capie, U. of Georgia, Athens, GA 30602
Valett Inventory of Critical Thinking Abilities (1981)	non-Piagetian	4-12	General	1	1	*	Open-ended	N.A.	Hand	Most	Some	Most	Most	Children complete tasks until child misses 4 out of 5 in a row	Academic Therapy Publications, 20 Commercial Blvd, Novato, CA 94947-6191
Understanding in Science (1979)	Piaget	7-9	Science	1	1	24	M.C., Short resp.	40 min.	Hand	Most	Some	Most	Some		Australian Council for Educational Research, Ltd., Frederick St., Hawthorne, Victoria 3122, Australia

ARTER HANDOUT

Instrument	Focus	Grades	Subject Specificity	No. Forms	No. Levels	No. Items	Item Type	Adm. Time	Scoring	Norms	Other Interp.	Reliability	Validity	Comments	Availability
CREATIVITY TESTS:															
Make a Tree (1976)	Divergent	preK-1	General	1	1	2	Open-ended	Not provided	Hand	Good	Some+	None	None	Part of the CIRCUS Test Battery. Lots of help with interpretation	CTB/McGraw Hill 2500 Garden Road Monterey, CA 93940
Pennsylvania Assessment of Creative Tendency (1968)	Affective Correlates of Creativity	4-9	General	2 short 2 long	1	19 short 39 long	Likert	Not provided	Hand	Fair	Some	Good	Some+	Long and short forms of the survey are available. Short forms should only be used experimentally	Tests in Microfiche (PATS09), ETS Test Collection Princeton, NJ 08541-0001
Possible Jobs (1963)	Divergent Thinking	6-12	General	2	1	3	Open-ended	10 min.	Hand	Fair	Some	Fair	Some+		Sheridan Psychological Services P.O. Box 6101 Orange, CA 92667
Seeing Problems (1969)	Sensitivity to problems	7-adult	General	2	1	6	Open-ended	7 min.	Hand	Fair	Some	Fair	Some+	Although this seems to represent the factor claimed, usefulness in all educational settings has not been demonstrated.	Sheridan Psychological Services P.O. Box 6101 Orange, CA 92667
Test of Creative Potential	Divergent thinking	2-12	General	2	1	3	Open-ended	30 min.	Hand	Fair	None	Fair Good	Some		Monitor, P.O. Box 2337 Hollywood, CA 90078
Test of Divergent Thinking (1980)	Divergent thinking	3-12	General	1	1	12	Open-ended	20-25 min.	Hand	Fair	Some	Poor	Some	Part of the Creativity Assessment Packet	DOK Publishers E. Aurora, NY 14052
Test of Divergent Feeling (1980)	Affect	3-12	General	1	1	30	MC	20-30 min.	Hand	Fair	Some	None	Some	Part of the Creativity Assessment Packet	DOK Publishers E. Aurora, NY 14052
Williams Scale (1980)	Creativity General	1-12	General	1	1	48	Check-list	30 min.	Hand	Fair	Some	None	Some	Part of the Creativity Assessment Packet	DOK Publishers E. Aurora, NY 14052
Thinking Creativity with Sounds and Words (1973)	Creativity-General	3-adult	General	2	2	8	Open-ended	30-35 min.	Hand	?*	?*	Interrater-Excellent	?*	Uses a recording for sounds. Some information in a technical manual which we did not get.*	Scholastic Testing Service, 480 Meyer Road Bensenville, IL 60106
Torrance Test of Creative Thinking (1970)	Divergent thinking	K-adult	Science	2	1	10	Open-ended	1 hr 45 min.	Hand	Good	Some	Good	Extensive	Based on a broad definition of the creative act.	Scholastic Testing Service, 480 Meyer Road Bensenville, IL 60106

*Technical manual not available

TEST	FOCUS	GRADES	SUBJECT AREAS	SCORE OBTAINED	AVAILABILITY
ACHIEVEMENT TESTS:					
Assessment of Reading Growth (1980)	Inferential Comprehension	3,7,11	Reading	Inferential comprehension score available	Jamestown Publishers P.O. Box 6743 Providence, RI 02940
California Achievement Test (1985)	Bloom's Taxonomy	K-12	Reading Language	HOTS score re-scored from other subtests by publisher	CTB/McGraw Hill 2500 Garden Road Monterey, CA 93940
Comprehensive Tests of Basic Skills (1981)	Bloom's Taxonomy	K-12	Reading, Math Language, Science, Social Studies, Ref. Skills	Items are cross-referenced to Bloom's Taxonomy. User must generate HOTS score.	CTB/McGraw Hill 2500 Garden Road Monterey, CA 93940
Iowa Test of Basic Skills (1985)	Various depending on subtest	K-12	Listening, Reading, Maps, Ref. materials	Same as above.	Riverside Pub. Co. 8020 Bryn Mawr Avenue Chicago, IL 60631
Metropolitan Achievement Tests (1985)	Bloom's Taxonomy	K-12	Reading, Math Science, Social Studies	HOTS score re-scored from other subtests by publisher	Psychological Corp. 555 Academic Court San Antonio, TX 78204
National Tests of Basic Skills (1985)	Inference Evaluation	PreK-adult	Reading	Inferential and evaluative comprehension aggregated by publisher; % correct on individual objectives.	American Testronics P.O. Box 2270 Iowa City, IA 52244
Reading Yardsticks	Interpretation Evaluation	K-8	Reading	Interpretive and evaluative reading available from publisher	Riverside Pub. Co. 8020 Bryn Mawr Avenue Chicago, IL 60631
Scan-Tron Reading Tests (1985)	Inferential Comprehension	3-8	Reading	Items are cross-referenced to skills. User must generate score.	SCAN-TRON Corporation Reading Test Division 2021 East Del Amo Bv. Rancho Dominiquez, CA 90220
SRA Achievement Series (1985)	Various, depending on subtest	K-12	Reading Math Social Studies Science	Items are cross-referenced to skills. User must generate score.	Science Research Associates, Inc. 155 Wacker Drive Chicago, IL 60606
Stanford Achievement Test (1981)	Using information	1-9	English Science Social Studies	Using information score, re-scored from other subtests by publisher.	Psychological Corp. 555 Academic Court San Antonio, TX 78204-0932
Stanford Test of Academic Skills (1982)	Using information	8-adult	English Science Social Studies	Using information score, re-scored from other subtests by publisher.	Psychological Corp. 555 Academic Court San Antonio, TX 78204-0932
Survey of Basic Skills	Various, depending on subtest	K-12	Listening, Reading, Math Social Studies Science	Users consult the skills profile report for each subtest	Science Research Associates, Inc. 155 Wacker Drive Chicago, IL 60606
The Three R's Tests (1982)	Logical relationship Literary analysis Author's purpose problem solving	K-12	Reading Language Arts Mathematics	Item analysis scores available from publisher	Riverside Pub. Co. 8420 Bryn Mawr Ave. Chicago, IL 60631

Interpretation of Table Codes

Norms (Value judgement implied)

None	No normative information is provided
Fair	Has some standards of comparison, e.g., means of research sample, decile norms or item statistics.
Good	Has norms based on a good sized sample or lots of other information.
Excellent	Has norms based on a national sample, and other information.

Other Interpretation (No value judgement as to the quality of the assistance is implied)

None	No help with interpretation provided.
Some	Has some help with interpreting scores, e.g., what the various scores mean.
Some+	Has information on what the scores mean and some help with use in instruction.
Extensive	Has extensive information on what the scores mean and how to use them in instruction.

Reliability (Value judgement implied)

None provided	No information was found.
Poor	All r's below .70
Fair	At least one reported r is greater than .70
Good	Total r is greater than .85; most subtests have r greater than .75.
Excellent	Several kinds reported; total score r is greater than .90; most subtest scores greater than .80

Validity (This describes the quantity of information available, not necessarily the extent to which the instrument is valid.)

No information	No information on validity is reported.
Some information	At least one activity related to validation is reported.
Some+ information	Validity was examined in several different ways.
Extensive information	Special effort was made to examine validity and there is a large research base on the instrument.

Checklist for Selecting a Higher Order Thinking Skills Test

I. Usefulness

A. Information Obtained

1. Do the stated uses of the instrument match up with what you want to use the information for?

2. Does the instrument or method measure the HOTS skills on which you want information?

3. Does the instrument assist with interpretation of results? Does it have criteria by which to judge results? This includes statements about what performance should be like at various grade levels. It could also include norms.

4. Is there information about how to use the results to plan instruction for students?

B. Logistics

1. Is the instrument or method easy to use?

2. Is it easy to score and interpret the results?

3. Is the length of time required to collect information acceptable?

C. Cost

1. Are costs within available resources? (Include costs of obtaining the instrument or method, training data collectors and collecting data.)

II. Technical Adequacy

A. Theoretical Basis

1. Do the supporting materials for the instrument or method present a clear definition of the aspects of HOTS that it claims to measure? Does the test manual discuss how this definition was developed and why the test has the content it has? Is evidence provided (based on research or theory) that the definition(s) and test content are reasonable?

B. Reliability

1. Was the instrument pilot tested?

2. Is there some measure of reliability available for the instrument?

a. For a structured-format test this includes at least item discriminations, internal consistency and test-retest reliabilities.

b. For an open-ended test this would include estimates of reliability of scoring such as interrater reliability.

c. If the results are going to be used to make important (and hard to reverse) decisions about individual students, reliability should be above .90. For group uses, or for educational decisions that are easily reversible, reliabilities should be above .75.

C. Validity: Is there evidence that the instrument measures what it claims to measure? Validity is in the relationship between the instrument and its use. There should be evidence that the instrument can be validly used for the purposes stated. For example, what evidence is there that the item types used measure the skill area?

1. For structured-format instruments an ideal set of validity studies would include:

a. The respondent understands what is being asked. Vocabulary or concepts unfamiliar to a group would make the instrument unusable for that group. This information would most likely be obtained by observing or interviewing students.

b. Right answers are only arrived at through the thinking process claimed to be measured not from clues or faulty assumptions. Likewise wrong answers are arrived at through faulty reasoning and not due to good reasoning based on a different philosophical orientation or experience level. This information would most likely be obtained by observing or interviewing students.

c. There is a moderate correlation with intelligence and achievement tests. Scores correlate with other validated tests claiming to measure the same thing.

d. There is a factor analysis done to show that the subscales do measure different things.

e. Groups that should be different in their scores are indeed different. This could include the ability of an instrument to differentiate between types of students.

f. The instrument measures changes or differences in HOTS after training designed to change HOTS.

g. There is a clear and frank discussion of the measurement issues involved including which aspects were investigated during the development process and which were not.

h. It is the opinion of knowledgeable judges that the instrument measures the HOTS aspects claimed.

i. For Piagetian instruments there is a high correlation between scores on the test and level of formal reasoning obtained from clinical interviews.

2. For open-ended instruments this would include:

a. The respondent understands what is being asked. Vocabulary or concepts unfamiliar to a group would make the instrument unusable for that group.

This information would most likely be obtained by observing or interviewing students.

b. There is a moderate correlation with intelligence and achievement tests. Scores from the instrument correlate with scores from other instruments claiming to measure the same thing.

c. Groups that should be different in their scores are indeed different. This could include the ability of an instrument to differentiate between types of students.

d. The instrument measures changes or differences in HOTS after training designed to change HOTS.

e. There is a clear and frank discussion of the measurement issues involved including which aspects were investigated during the development process and which were not.

f. It is the opinion of knowledgeable judges that the instrument measures the HOTS aspects claimed.

Diagnosing Thinking Deficiencies in the Classroom

Selma Wassermann.

Dr. Wassermann began by making a connection between the relatively new fields of teacher testing and assessing thinking skills. She feels that both are attempting to measure acts which are:

-complex, sophisticated and interrelated;
-observable and intuitive;
-creative;
-inconsistent;
-independent and interrelated;
-ambiguous and nonlinear.

The hazards of assessing students thinking skills with paper and pencil tests are that the tests:

-drive the curriculum;
-produce linear thinkers;
-lose intuitive, sloppy acts of geniuses;
-have only a single score;
-restrict thinking with forced choices;
-label and categorize people;
-take away important teacher judgments (observation);
-discriminate against slower thinkers.

In lieu of paper and pencil tests, Wassermann suggests we use a variety of means to assess thinking: behavioral scales, performance on tasks which use a variety of skills, and teacher observation. The problem with these means is that they are high inference and therefore "soft" data but done properly can be highly effective and relevant.

Observation, performance and behavioral measures:

-contribute to our overall understanding of the student;
-provide the most valuable data;
-provide long-range data;
-look at thinking applied in many contexts as well as when it is not applied;
-can be used for targeting thinking deficits in individual student instructional plans.

The remainder of the workshop was a hands on application of Wassermann's "Teacher Rating Instrument" which has eight student "types" displaying different thinking deficits. Many of these deficits appear to be affective and would be improved by positive self-esteem. Wassermann would disagree. She maintains that these are cognitive deficits. The instrument and two case studies are attached.

Task 1:

(1) Work in pairs.

(2) Read through the eight behavior profiles in the Teacher-Rating Instrument. Talk with each other about any students you have observed who exhibit these patterns.

(3) Then, read the brief Case Study of Bob.

(4) Talk with each other about what you perceive to be important behavioral indicators of Bob's deficits in his thinking capability.

(5) Use the Teacher--Rating Instrument and decide on how you will assess Bob's behavior.

(6) Follow the same procedures with the Case Study of Eddy.

Task 2:

(1) Work in pairs.

(2) Read through the 33 behavioral indicators on the Profile of Student Performance. Talk with each other about students you have observed who exhibit these characteristics in their academic task work.

(3) Then, re-read the Case Study of Bob. Discuss what you perceive to be any observable positive or negative indicators in his performance on the Comparing assignment.

(4) Using the Profile of Student Performance, decide on an appropriate assessment of Bob's ability to compare.

(5) Follow the same procedure with the Case Study of Eddie.

TEACHER RATING INSTRUMENT

Selma Wassermann
Faculty of Education
Simon Fraser University
©1986

Notes to the teacher:

The teacher rating instrument helps you identify students who are exhibiting behaviors associated with deficits in higher-order thinking capabilities. Used in pre- and post-assessments, it also provides means for examining positive gains in those behaviors associated with thinking.

Read each profile. In section A, after each profile, write the names of those students whose behavior you believe to be clearly described in the profile. In section B, after each profile, write the names of those students whose behavior may possibly reflect that described in the profile (i.e., you are not certain; there is some element of doubt).

1. Profile of the very impulsive student.

These students typically act without thinking. When a problem or activity is introduced, these students leap into action first. They don't seem to have a plan, nor do they consider alternatives. The mode of operation is *doing* -- and sitting down to "think things out" does not seem to be their pattern of behavior. Impulsive students do not engage in reflection.

Associated with impulsiveness is the idea of rapid and random movements, of acts directed by little more than whim or caprice. Action itself seems to be much more important than thinking about possible modes of action. To be up and doing is so prominent in the behavior of these students that the purpose or goal of the doing is neglected.

A. This pattern of behavior sounds very much like:

1 ..... 3 .....

2 ..... 4 .....

B. This pattern of behavior sometimes describes:

1 ..... 3 .....

2 ..... 4 .....

2. Profile of the very dependent student.

These students typically want help with practically everything they undertake. They find it hard to begin work without asking for help in getting started. Once they have begun, help is again requested. Often such students say, "I'm stuck!" or "What shall I do now?" or "I don't know what I'm supposed to do."

These students' insistent calling on the teacher for help is a strong characteristic of their inability to carry out tasks independently. When the teacher's help is not available, these students may just sit there and do nothing until help is forthcoming.

A. This pattern of behavior sounds very much like:

1 ..... 3 .....

2 ..... 4 .....

B. This pattern of behavior sometimes describes:

1 ..... 3 .....

2 ..... 4 .....

3. Profile of the loud, dogmatic and overly assertive student.

These students seem to have "all the answers," and they are unyielding in their conviction that they are *right.* Such students reject discrepant data; their minds are made up. These students stick to their positions regardless of the facts. There is a quality of inflexibility about the way they think. We generally consider them as intemperate, unreasonable and insensitive to the feelings of others. They don't listen to alternatives. In fact, they seem reluctant to acknowledge that there are alternatives.

These students are apt to impugn the motives of those who oppose their views. Their language is frequently studded with extreme words, such as *always, everybody, nobody, never*. They may generalize loosely about races or nations or religious groupings. We sometimes think of them as authoritarian; that they try to dominate other people; that they are rash in their judgments, close-minded, intemperate, or unreasonable in presenting their ideas.

A. This pattern of behavior sounds very much like:

1 ..... 3 .....

2 ..... 4 .....

B. This pattern of behavior sometimes sounds like:

1 ..... 3 .....

2 ..... 4 .....

4. Profile of the rigid, in-a-rut student.

These students typically want to stick to doing things in the same old ways. They don't like new or different ways of doing things. The fact that a problem or task is new and calls for new procedures doesn't make a difference. These students try to force old methods onto new problems. When that procedure doesn't work, these students complain that the problem is at fault.

These students are most comfortable when carrying out routines. They can learn lessons and formulas, but they have great difficulty in applying the principles of what has been learned to new situations.

A. This pattern of behavior sounds very much like:

1 ..... 3 .....

2 ..... 4 .....

B. This pattern of behavior sometimes describes:

1 ..... 3 .....

2 ..... 4 .....

5. Profile of the student who misses the meaning.

These students don't seem to understand. They miss the point -- of a lesson, an assignment, a story, a joke. Such a student might say, "I don't get it." We think of these students as people who "don't listen" and who don't pay attention. Actually, the problem is that they are not able to interpret data intelligently or extract meaning from their experiences. It is as if their ability to process data has been seriously impaired.

A. This pattern of behavior sounds very much like:

1 ..... 3 .....

2 ..... 4 .....

B. This pattern of behavior sometimes describes:

1 ..... 3 .....

2 ..... 4 .....

6. Profile of the student who can't concentrate.

The dominant characteristic of these students is their tendency to use means that are inconsistent with or inappropriate to the ends they seek. It's not that these students don't have any ideas. They do have ideas about what they want to do, but the paths that they take to arrive at their goals may be silly or illogical, impractical, or even irrational. There seems to be an absence of awareness in choosing these paths. Their choices do not reveal a great deal of thinking about the connections between means and ends, but rather suggest an indiscriminate and random selection of means. Perhaps that accounts for the fact that the goals these students set are seldom realized.

A. This pattern of behavior sounds very much like:

1 ..... 3 .....

2 ..... 4 .....

B. This pattern of behavior sometimes describes:

1 ..... 3 .....

2 ..... 4 .....

7. Profile of the underconfident student.

These students lack confidence in expressing their thoughts. During class or group discussions, such students rarely volunteer information, not because they have nothing to say, but because of fear of exposing what they are thinking to public scrutiny. At the end of class, these students might say privately to the teacher that they did have an idea, but were concerned about how it might have been received by the group. Underconfident students are not necessarily shy; they rather lack confidence in their ability to think and are fearful about exposing their thoughts to possible criticism.

A. This pattern of behavior sounds very much like:

1 ..... 3 .....

2 ..... 4 .....

B. This pattern of behavior sometimes describes:

1 ..... 3 .....

2 ..... 4 .....

8. Profile of the student who actively resists thinking.

These students typically scorn thinking as a preferred mode of operation. They believe that action is more important and that thinking is for the "intellectuals." Classmates who are seen as intellectually advanced are held in contempt. When asked to do some thinking, these students reject the process. "It's the teacher's job to tell us what to do" may be the response these students give. Such students see their function as acting rather than reflecting. They make strong value judgments about "people who think," and those judgments are clearly negative.

A. This pattern of behavior sounds very much like:

1 ..... 3 .....

2 ..... 4 .....

B. This pattern of behavior sometimes describes:

1 ..... 3 .....

2 ..... 4 .....

CASE STUDY OF BOB (A sixth grade boy)

Bob's cumulative school records showed that he had difficulties in school from first grade on. His first grade teacher wrote that his work habits and general attitude were poor. "Although he has done adequate work, I feel he could do better if he would only apply himself." Notes from his second grade teacher read, "I am constantly after him to get on with his work. His work is not as good as it could be. He just isn't trying." From the third grade teacher: "While Robert appears to have the ability to do better, he is functioning on a below average level in all of his subjects. He is a big disturbance in class." Other teachers noted is "wandering behavior, his lack of preparation, his tendency to procrastinate; that he was the last to finish his work." The fourth grade teacher wrote that "he finds it difficult to stay put for long; that he lacked the ability to stay on task."

His sixth grade teacher noted: "He needs close supervision and structuring. He seems unable to function independently. He seems unable to put his mind to a task; instead of engaging, he avoids. He seems to need to conform and becomes very agitated when he is asked to think things out for himself. He expresses deeply held racist views and is very adamant about his point of view and rarely allows that another opinion has validity."

The following are some selected events that are typical of his classroom performance. (It should be noted that Bob was considered to be of "normal" intelligence; that he had no other discernable "learning disability;" and that he suffered from no observable emotional disturbance.)

** When he was asked to do some planning for his social studies project, he said (to the teacher): "You're the teacher. It's your job to tell me what to do."

** When he observed that the classroom sink had become clogged, he offered: "You know why this thing's clogged -- the plumbing is cheap."

** After the teacher had been absent for a day, he said, "Don't be absent anymore. I need you here. I need you to tell me what to do."

** In a discussion on the Hindus and Muslims, he said, "What's this about them believing that cows are sacred? Cows running loose and being worshipped? Isn't that pretty stupid?"

** Bob: "You know, we eat protozoa. Meat has cells and we eat meat; so we eat protozoa."

When asked to compare a news article with an editorial (he was given copies of each), he wrote:

"My thought. As you can see, there is much difference between the articles."

CASE STUDY OF EDDIE (A sixth grade boy)

Eddie is an attractive boy; gregarious and appealing in many ways. Of "normal" intelligence, he has consistently been categorized as "under achieving." He is physically healthy and school records show no discernable evidence of emotional instability.

From early first grade experiences, Eddie was described as "quite immature; noisy, restless and annoying to other children." His third grade teacher noted that he tended to rush through his assignments, making careless, thoughtless errors; that in his work, speed rather than accuracy seems to be his goal." The fourth grade teacher surveyed her year with him by writing, "He is easily diverted from any task. As a result, he jumps from activity to activity without completing assigned material. He seems to need constant attention and constant encouragement. He cannot concentrate. I have tried to point out to him that we cannot accomplish a great deal by doing three or four things at one time; but this seems difficult for him to understand. He does not follow directions, written or otherwise. Naturally all this behavior directly affects his academic performance." Eddie's fifth grade teacher indicated similar problems: "Eddie doesn't listen. He has difficulty following directions. This causes him to frequently do the wrong work. Often he misinterprets what has been said." The sixth grade teacher noted: He has a great deal of difficulty doing independent work. His work is haphazard and impatient. He is most comfortable when following routines; if something new is proposed, he rejects it and asks for 'doing it the way we used to.'"

The following are some selected events that are typical of his classroom performance:

** Eddie (to his teacher): Sometimes when you (teacher) say it's all right to do our work in pencil, I start to use a pencil. Then I look at the guy next to me and he is using a pen. The guy next to him is using a pen. So I rip up my paper and start all over again with a pen.

** Eddie: "I have a good idea for a science project. I want to do a project on static electricity with balloons. I did that last year, too."
Teacher: "How come you want to do it again?"
Eddie: "It was a good project last year and I want to do it again."

** Eddie: "I don't understand. What are we supposed to do here?"

** Eddie: "Just tell me what I'm supposed to do. I'll do it."

** Eddie: "I like to do things the way I've done them before. I don't like new ways of doing things.'

** Eddie: "When I get stuck, I look around for somebody to help me. I have a hard time getting started, because I'm not sure about what I'm supposed to do."

PUPIL'S NAME_____

DATE OF ASSESSMENT (1)_____

(2)_____

THINKING SKILLS DEVELOPMENT PROGRAM

PROFILE OF STUDENT PERFORMANCE

("How Well Does Sammy Think?")

©Selma Wassermann
Faculty of Education
Simon Fraser University
1984

WASSERMAN
HANDOUT

This 33 paired-item rating scale is designed to help the teacher diagnose a student's performance on tasks which require thinking at higher cognitive levels.

Each of the paired items reflects a "positive" and a "negative" view of pupil behavior in the performance of thinking tasks. Since a student's performance may vary, your assessment on each of the items should reflect your impression of a student's overall functioning over a period of time, rather than on a single task. In making your diagnosis, use the following scale as a guide:

Positive behavior — Thinking-related behavior

+4	+3	+2	+1	-1	-2	-3	-4
This behavior is seen almost all the time	This behavior is frequently seen	This behavior is seen sometimes	This behavior is rarely seen	This behavior is rarely seen	This behavior is seen sometimes	This behavior is frequently seen	This behavior is seen almost all the time

What do the ratings tell us?

Ratings which are consistently on the "minus" side of the scale would reveal that a student is having considerable difficulty with thinking skills. A student with a high "minus" score would be manifesting extreme "thinking-related behaviors" -- that is, behaviors associated with inadequate cognitive functioning. Higher negative ratings indicate that a student needs much more work on the development of thinking skills.

Ratings which are consistently on the "plus" side of the scale point to students who are exhibiting successful thinking skills development. Higher positive ratings point out the gifted and talented thinkers, who are likely to benefit enormously from additional work on thinking tasks.

This diagnostic assessment should provide the teacher with an analysis of individual pupils' strengths and weaknesses in performing thinking tasks. The most important use to be made of this instrument is to enable the teacher to use each diagnosis in helping each student to further develop his/her thinking capabilities.

2.5
2.2
2.0
2.8
3.2
3.6
4.0
100 mm
1.0
1.1
1.25
1.4
1.6
2.5
2.2
2.0
1.8
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890
1.0 mm
1.5 mm
2.0 mm
A5

A. PUPIL PERFORMANCE ON OBSERVING TASKS

Ratings	Positive behaviors	Thinking-related behaviors	Ratings
	1. To what extent does this student show that his/her observations reflect careful and thoughtful attention to substance and detail?	1a. To what extent are the student's observations hasty and manifest of "insufficient" thought?	
	2. To what extent are the observations accurate?	2a. To what extent does the student go beyond the data and make unsupported assumptions in his/her observations?	
	3. To what extent is the student discriminating in his/her observations?	3a. To what extent is he/she satisfied with superficial treatment of the material?	

Teacher's diagnostic observations of pupil's observing skills:

WASSERMAN HANDOUT

B. PUPIL PERFORMANCE ON CLASSIFYING TASKS

Ratings	Positive behaviors	Thinking-related behaviors	Ratings
	4. To what extent does this student show an ability to create groups for the categorization of objects?	4a. To what extent does the student get "stuck" in the formation of groups?	
	5. To what extent are the student's groups related to some common principle?	5a. To what extent are the groups random, and unrelated to each other?	
	6. To what extent is he/she able to place objects into the groups where they belong?	6a. To what extent is he/she unable to show discrimination in the sorting and classifying of objects?	

Teacher's diagnostic observations of pupil's classifying skills:

WASSERMAN HANDOUT

C. PUPIL PERFORMANCE ON COMPARING TASKS

Ratings	Positive behaviors	Thinking-related behaviors	Ratings
	7. To what extent do the comparisons show careful and thoughtful observations of substance and detail?	7a. To what extent are comparisons hasty and reflective of superficiality?	
	8. To what extent do the comparisons enable the student to arrive at deeper meanings about the items being compared?	8a. To what extent are the comparisons reflective of only the surface aspects of the items?	
	9. To what extent are the comparisons comprehensive in scope?	9a. To what extent are the comparisons reflective of "insufficient" thought?	

Teacher's diagnostic observations of pupil's comparing skills:

WASSERMAN HANDOUT

D. PUPIL PERFORMANCE ON LOOKING FOR ASSUMPTIONS TASKS

Ratings	Positive behaviors	Thinking-related behaviors	Ratings
	10. To what extent does the student show that he/she is aware of the differences between fact and assumption?	10a. To what extent is the student "stuck" and unable to comprehend the concept of assumptions?	
	11. To what extent is the student able to identify several assumptions underlying an event?	11a. To what extent is he/she satisfied with the finding of a single assumption?	
	12. To what extent does the student place value on the importance of discriminating between assumption and fact?	12a. To what extent does he/she consider the identification of assumptions inconsequential?	

94

Teacher's diagnostic observations of pupil's skills in looking for assumptions:

WASSERMAN HANDOUT

E. PUPIL PERFORMANCE ON HYPOTHESIZING TASKS:

Ratings	Positive behavior	Thinking-related behavior	Ratings
	13. To what extent is the student able to come up with several thoughtful and appropriate hypotheses for problem situations?	13a. To what extent is he/she satisfied with a single response?	
	14. To what extent is the student capable of suspending judgment and keeping an open mind in the examination of alternative hypotheses?	14a. To what extent is he/she dogmatic about proposing solutions?	
	15. To what extent is the student comfortable with the cognitive dis-sonanace underlying the tentative and provisional aspects of hypothesizing?	15a. To what extent does this kind of ambiguity make the student uncomfortable?	
	16. To what extent is he/she capable of performing hypothesizing activities using his/her own thinking?	16a. To what extent is he/she dependent upon the teacher's help in carrying out the tasks?	

95

Teacher's diagnostic observations of pupil's skills in hypothesizing:

WASSERMAN HANDOUT

F. PUPIL PERFORMANCE ON IMAGINING TASKS:

Ratings	Positive behavior	Thinking-related behavior	Ratings
	17. To what extent is the student able to create new and original schemes?	17a. To what extent does he/she tend to repeat and draw upon forms which have been done many times before?	
	18. To what extent does the student enjoy creating and imagining?	18a. To what extent is this an intimidating activity for this student?	
96	19. To what extent does the student's imagining activities show a talent for creativeness, inventiveness and originality?	19a. To what extent are this student's imagining activities routine, standard and lacking in creative flair?	

Teacher's diagnostic observations of pupil's imagining skills:

G. PUPIL PERFORMANCE ON DECISION-MAKING SKILLS

Ratings	Positive behavior	Thinking-related behavior	Ratings
	20. To what extent is this student able to think through a situation and arrive at a decision?	21a. To what extent does this student get "stuck" and unable to come to his/her own decision?	
	21. To what extent is he/she able to take ownership for the decision made?	21a. To what extent does this student avoid taking responsibility for his/her decision?	
	22. To what extent is he/she "in touch" with the values he/she holds which guide that decision?	22a. To what extent does this student's decision seem to be made up on the spur of the moment and not guided by a clear and consistent value system?	

97

Teacher's diagnostic observations of pupil's decision-making skills:

WASSERMAN HANDOUT

H. PUPIL PERFORMANCE ON PROBLEM SOLVING SKILLS:

Ratings	Positive behavior	Thinking-related behavior	Ratings
	23. To what extent is this student able to function on problem solving tasks on his/her own?	23a. To what extent is he/she dependent upon others for help with each step of the problem?	
	24. To what extent is he/she able to gather data and process the data?	24a. To what extent does he/she get stuck and bogged down, and unable to complete the process?	
	25. To what extent is he/she able to come up with possible solutions that reflect the data gathered?	25a. To what extent are the solutions "made up" or guesses?	
86	26. To what extent is he/she able to come up with several possible solutions?	26a. To what extent is he/she looking for just a single answer?	

Teacher's diagnostic observations of pupil's problem-solving skills:

WASSERMAN HANDOUT

I. ADDITIONAL PERFORMANCE CRITERIA FOR ALL THINKING TASKS:

Ratings	Positive behavior	Thinking-related behavior	Ratings
	27. To what extent does the student give his/her full attention to the task of thinking?	27a. To what extent does the student give only token attention to the tasks?	
	28. To what extent is the student able to take cognitive risks and give responses that show originality and creativity?	28a. To what extent are the responses more standard, and more traditional in their approach?	
	29. To what extent does the pupil show that he/she values the process of thinking?	29a. To what extent does the pupil pay more attention to the product -- i.e., getting the "right" answer?	
	30. To what extent is the pupil able to carry the tasks to completion?	30a. To what extent does he/she leave the tasks unfinished?	
	31. To what extent does the pupil show an understanding of the process of the operation behind each activity?	31a. To what extent is the pupil responding in a rote, formula-like way?	
	32. To what extent does the pupil show his/her ability to deal with situations where there are many acceptable answers?	32a. To what extent does the pupil need reassurance that the work is "right" and "good?"	
	33. To what extent does the student show a high level of engagement of thought on the tasks?	33a. To what extent does he/she avoid the thinking tasks because thinking is too hard?	

Teacher's diagnostic observations of overall pupil performance:

WASSERMAN HANDOUT

Assessing Creativity

Anita Halstead

Creativity is a relatively new concept that Ms. Halstead defines as "the ability to go beyond perceived patterns and rules and have value." There is general agreement that the four skills that measure creativity are:

- Fluency: ability to generate a lot of ideas;
- Flexibility: ability to take many different viewpoints;
- Elaboration: ability to extend an idea, add, refine it so it is better understood,
- Originality: ability to produce statistically unique ideas.

In addition there are issues of: sensitivity to problems, leadership, and humor.

There are many paper and pencil creativity tests available but these kind of tests do not measure the "essence" of creativity. Halstead favors informal assessments, followed up with positive reinforcement (for those who do favorably) and action. Act on the results rather than file away for the future. Tests can give people knowledge about themselves. Most teachers and supervisors cannot predict who will do well on a creativity measure.

Divergent questions can informally assess creative thinking:

- Generate ideas: both within and outside of content areas;
- Viewpoint: look at something from a different perspective ("how would an ant see this?");
- Involvement: affective;
- Conscious self-deceit: (Hypothetical wishing. If I had...)
- Forced association: how are two unlike things similar? How are two similar things different?
- Reorganization model: looking at something in a different way;

Ms. Halstead related several informal performance measures for assessing left and right brain dominance as well as asking us to solve a problem which we scored for fluency, flexibility, elaboration and originality.

She maintains that it is possible to train people to think creatively by making them aware of the creative process. Children and adults are often the best judge of their own creativity. But when making formal assessments of children, it is important to use multiple sources of data and to be cognizant of children's developmental stages which are:

- Dreamer: 18 months-2 years; fantasy and nonverbal activity;
- Poet: 3-6 years; rhyming, nonsense language (Dr. Seuss); playing with words
- Inventor: 6-8 years; takes things apart; inventive; mixes materials; wants to know how things work;
- Poet: 10-12 years and in 20's; plays with semantics; journal, diary writing;
- Inventor: 12-16 years and 35-40 years; challenging and wondering why things occur.

In general, a creative child has these qualities: IQ range of 118-130, controlling, dominant, unusual interests, high energy, adventuresome, temperamental, egocentric and disorganized.

Assessing Higher Order Thinking Skills in Hawaii: One State's Experience

Judith A. Arter and Lynde Paule

The purpose of this session was to describe some work in progress that NWREL is doing under contract to the Hawaii State Department of Education. The project is to develop several sets of questions designed to measure various aspects of the State's learning objectives for students, called FPOs.

One of the eight areas to be assessed is "develop decision-making and problem-solving skills." The seven subskills listed by Hawaii for this objective are shown on page 1 of the presenters' handouts. In order to clarify the subskills to provide guidance for item development, NWREL staff listed various item types that could be used to address each subskill.

About 40 items were written to cover these skills. Dr. Arter related that the items were written with several goals in mind. First, the authors tried to make the situations in the test reflect real-life situations. Second, multiple questions were written for each situation so that the situation could reflect several pieces of information and so that the situation could be examined from several perspectives. After review by Hawaii State Department staff, 24 were selected for pilot testing.

The procedure used for the pilot test, and the rationale for this procedure, are provided on pages 2-5 of the presenters' handouts. The pilot test was designed to address some of the validity issues that surround multiple-choice tests of HOTS. In a multiple-choice test the right answer is a proxy for the skill that the item is supposed to measure. That is, if a student gets the item right then the assumption is that the student has demonstrated the skill. Similarly, if the student gets the item wrong then the assumption is that the student has not demonstrated the skill -- in this case good thinking. The general procedure was to have a group of students take the test orally -- read everything out loud, indicate words and passages that were not understood and provide the rationale for the answer they choose. If students typically get a question right or wrong for other than the intended reasons then the question will be altered or dropped.

Twenty-eight students were tested in this manner. Each tape is from 70 to 120 minutes long. Two raters reviewed the tapes. The scoring system used to rate the rationale for answers is provided on page 6 of the handouts. In addition to scoring all distractors and right answers in this manner, the raters made lots of notes on each student in an attempt to adequately capture thinking processes.

Although the data are still being analyzed, Dr. Paule related that the results relate both to items and to students. Problems with items included:

1. Differences in the general knowledge that different students bring to the task. This influences several items.

2. Differences in cultural values that affected one or two of the items. Specifically, there was an item than required the student to select the best solution to a problem. One student based his rationale on the value that agreement between people is more important than anything else.

3. Many of the passages and questions appeared to be at too difficult a reading level. Sometimes this indicated a need to change wording. Other times it indicated student deficiencies -- for example, knowledge of critical thinking terminology such as "criteria," "assumption," and "unbiased."

4. A few times students came up with good reasons why the response we thought was right was not right, and why a distractor might be the better choice.

The process may also have uncovered some deficiencies in student thinking patterns. "May have" is important here since the sample is small, and because sometimes when students had extreme difficulty in providing rationales for answers it was hard to tell whether there was a problem with the question or a problem with the student thinking pattern. Some issues with respect to students that should be followed up include the following. In the presentation, each of these points was illustrated with a portion of one of the student tapes.

1. Some students seemed to have difficulty putting themselves in someone else's shoes. Some questions required the student to pretend that they had a particular list of concerns and respond from that viewpoint. Students would often ignore this task and answer from their own opinion.

2. Students often did not answer the question or refer back to the passage to answer the question. Rather they would answer each option as being a separate true/false question or drift off into wild flights of fancy as to why a choice is right or wrong.

3. Many students seemed to have difficulty articulating why they choose the answer they did.

4. As mentioned above, many students had a lot of trouble with thinking skills terminology.

FPO III Content Description

A1. Identifies, clarifies, and states a problem and develops criteria for examining alternatives in solving the problem.

1.	Identify central issue or problem	Airbags, Pemba, Gorge, Cartoon
2.	Restate problem or paraphrase	
3.	Compare similarities, develop analogies	
4.	Develop criteria for judging	Drugs
5.	Design an experiment	Grades, Detergent, Lightbulbs

A2. Gathers information from various sources and analyzes and organizes the information to facilitate the formulation of alternatives.

1.	What other information is needed	Pemba, Airbags, Detergent, Lightbulbs
2.	What information is relevant/irrelevant	Gorge, Airbags
3.	Where to get information	Pemba
4.	Fact v. opinion/biased v. unbiased sources	Airbags
5.	Checking the consistency of information	Drugs
6.	Identify assumptions	Airbags, Detergent
7.	Appraising observations	Gorge
8.	Which two values are in conflict	

A3. Formulates hypotheses about a problem based on available information.

1.	Develop hypothesis	Detergent
2.	Formulate questions that lead to deeper understanding	

A4. Applies the criteria established to select an alternative.
A5. Evaluates the alternative selected for its effectiveness.

1.	Which proposed solution might be best	Pemba
2.	Applying criteria	Airbags

A6. Draws conclusions or generalizations based on the alternatives or hypotheses and related information.

1.	Inferences and deductions	Gorge
2.	Logical syllogisms	Baboons
3.	Recognizing adequacy of data	
4.	What information supports a conclusion	Grades, Gorge
5.	Cause and effect	
6.	Interpreting the results of an experiment	Detergent

A7. Validates and reports the conclusions and modifications, if any.

1.	What to do to validate choice	Pemba, Airbags
2.	Probable consequences	
3.	How to report results	Detergent
4.	Expanding results	Detergent

FPO III Pilot Test Procedure

Purpose

FPO III questions are intended to measure higher-order thinking skills. The purpose of the pilot test is to try to determine whether they are fulfilling this function. Essentially our position is the same as Norris and King (1984):

> "The view guiding this study is that validity investigations of ability tests involve three steps: (i) finding out whether examinees understand the tasks on the test in the intended way; (ii) given that the appropriate understanding is there, determining whether examinees use appropriate approaches to complete the required tasks; and (iii) if number of correct answers is to be used as the indicator of examinees' ability, determining whether the keyed answers are justified given appropriate strategies for arriving at answers." (p. 38)

> "The general principle of ability test validation underlying this study is that ability tests are valid to the extent that good thinking leads to good performance on the test and poor thinking leads to poor performance. The attempt is thus to explain performance in terms of thinking, and to do this there must be a description of the thinking processes which lead to performance. . . For a test to be suitably valid, there must be at least an overall tendency for good and poor thinking to be linked to good and poor performance respectively." (p. 39)

Thus, if we want to show that the questions measuring FPO III are really measures of higher-order thinking we need to find out the thinking that goes into arriving at answers. To do this we will ask students to "think out loud" as they complete the test. The role of the interviewer is to prompt the examinee to provide the rationale for both the answer he or she chose and for the answers that he or she did not choose.

Materials Needed

Instructions to students
Student form of the test (reusable)
Separate answer sheet on which students will mark their answers
One copy of the interviewer form of the test for each examinee
Tape recorder
One 90-minute cassette tape for each examinee
3-4 pencils
Reliable clock or watch
Clipboard

Overview of Procedure

The interviewer will introduce the task to the student (see the next section) and will distribute testing materials. The student will get the student form of the test and will mark his or her answers on the separate answer sheet. The interviewer will make sure that students are providing the rationale for their choices through probes and will make notes on the interviewer form as they go. The taped version will be used to clarify interviewer notes as needed. The interviewer will also circle any words that the student does not understand and provide the meanings for these words. Also, if the student does not understand a question, circle it and provide a clarification (without giving away the answer).

Questioning Strategy. Conduct the interview in a non-leading fashion. We wish to influence students' thinking as little as possible. When the student tells you which answer he or she thinks is correct do not let the student know if it is right or wrong. Do not point out flaws in the students' reasoning. You will ask questions mainly to lead the student to clarify why the student chose the answer he or she did.

- Ask the rationale for the right answer. Then ask why the student thought each of the other choices were wrong.

- Sometimes the reasons given for choices are not reasons. They may repeat the choice, for example. Ideally, we need to know the criteria by which a student judged answers, and why the choices did or did not satisfy the criteria.

  If the student does give a criterion for choosing and if the student makes clear why the choice fulfills that criterion and the other choices do not, go on to the next question.

  If the student identifies a criterion but does not state why the choice fulfills that criterion ask "What difference does that make to your thinking?"

  If the student does not identify a criterion or state why a choice fulfills a purpose (e.g., the student repeats the choice with little or no elaboration), ask "Could you explain a little more what makes you choose one more than the other?" For example, in question one the student says -- "This is the right answer because they need to see what the benefits from the project will be and what its gonna do on the environment." Ask -- "Why is this most important issue to decide?" Or ask "Could you explain a little more what makes you choose this more than the other answers?"

Interruptions To the Student's Narrative. When students are reading items interrupt only to clarify meanings. When students are telling you the reasons for their choices and non-choices wait until they are finished talking. Then ask for any clarifications.

Cautions.

1. Do not rush the interview by beginning to speak imediately after the examinee stops speaking. Wait for a few seconds for the examinee to continue.

2. Do not cut off examinee's reasoning by signalling that enough has been said, even though many examinnes will appear by the tone of their voices to seek such signals.

3. Do not endorse or criticize examinees' fact finding or reason giving.

Instructions To Students

A. Introducing the student to the task

There is no set script for introducing students to the task, but be sure to cover the following points. Keep talking until the student seems to be relaxed and ready to begin.

1. Inform the student of our purpose -- to attempt to develop the best test we can of people's ability to reason out the answers to questions.

2. Inform the student of his or her role -- to give us information about how people think when they take our test so that we can change the test where changes need to be made.

3. Inform the student that we are interested in finding out about the test and not about the person taking it, so there is no reason to feel any stress or pressure.

4. Inform the student that we want to find out as best we can what he or she is thinking while choosing answers to the questions, and that to do this we will be asking some questions as the test is taken.

5. Tell the student that we are taping the session so that if the notes the interviewer is taking are incomplete we can refer back to the tape.

6. Ask the student the questions on the Interviewer Form Cover Sheet -- name, major and GPA.

B. The tape recorder

1. Label the tape with the student's name and start the recorder.

C. Beginning the test

1. Read these instructions to the student:

"Read each question out loud. Tell me any words or questions that you don't understand. We need to know these so that they can be changed. I will tell you the meaning of the word or the question. Then pick the answer you feel is best and tell me why you chose that answer. Also tell me why you didn't choose the other answers."

D. After the test

1. Ask the student the questions on the last page of the Interviewer's Form of the Test.

2. Fill out the postinterview questions on the Interviewer's Form cover sheet.

CODES USED FOR SCORING STUDENTS' RESPONSES TO HOTS ITEMS:

0: No reason given/guessing/question too hard/ interviewer led student to answer

1: Good reason for why R is right

2: Good reason why a D is right

3: Good reason why a D is wrong

4: Good reason why R is wrong

5: Bad reason why R is right

6: Bad reason why a D is right

7: Bad reason why a D is wrong

8: Bad reason why the R is wrong

Test Publisher's Panel

Catherine Ross representing Academic Therapy, Gale H. Roid representing Psychological Corporation and Margaret Hill representing SRA
Moderator: Ken Bumgarner.

Gale Roid began the presentation by giving an historical perspective to the increased demand for teaching and assessing higher order thinking. During the 70s and 80s students began showing unexpected lower performances in inferential thinking. An analysis of NAEP items also showed less student competence with inferential items. He hypothesized that the back to basics approach in education may have contributed to the decline.

Dr. Roid asserted that the MAT 6 is the first battery which tried to separate a HOT index from an achievement battery. He feels that the MAT 6 and the SAT 7 Plus are major test batteries which measure HOTS across the curriculum. Other tests, such as the Watson Glaser, are tests to appraise specific elements of HOTS.

The test to use should be determined by:

1. The instructional approach and
2. The purpose of testing.

Margaret Hill discussed the new SRA Survey of Basic Skills which will have an individual HOTS score and national norms generated from 60 items imbedded within the test. The test also gives a client the possibility of customizing the test with their own items.

The HOTS questions in the SRA are included in the content areas. Dr. Hill reported that, in the reading section, 3/4 of the items deal with analytical or inferential information. The Language Arts items address the various tasks required in writing. The math items are non-routine items which include extraneous or insufficient information and which may require estimation. In both Science and Social Science, the students are assumed to have some knowledge and HOTS are tested within the content area. All HOTS items are "trying to press students at higher levels of thinking."

Catherine Ross began by discussing the behaviors various thinkers consider as important in writing HOTS test items. The views of Lippman, Ennis, Bloom and Costa were briefly discussed.

The Ross Test (designed for grades 4-6) includes items to assess specific skills of HOT with 8 test sections. These are tied to Bloom's Taxonomy.

1. Analogies,
2. Deductive Reasoning,
3. Missing Premises,
4. Abstract Relations,
5. Sequential Syntheses,
6. Questioning Strategies,
7. Analysis of Relevant and Irrelevant information,
8. Analysis of Attributes (a concept formation activity)

The Ross test is not intended to be comprehensive, but is planned to be a verbal expression of student HOT skill.

Ken Bumgarner summarized by saying that the presentations were a reminder that we can gain a great deal of knowledge about student HOTS from paper and pencil tests if items are creatively written.

Higher Order Thinking Skills and Item Banking

Richard Naccarato and Ray K. Miller

The first portion of this session was devoted to summarizing the results of an Item Bank survey which was conducted by NWREL in 1984. It was pointed out that *very few* of the banks surveyed had categorized their items by cognitive level, in order to allow selecting items which measured higher-order thinking skills. Some advantages/disadvantages of categorizing items by certain of these characteristics would be:

Advantages:

being able to select items to form a test of primarily "HOTS" knowledge

equating or adjusting difficulty levels of instruments based upon HOTS skills levels

Disadvantages:

HOTS taxonomical classifications for a particular item may spread across numerous categories

the item difficulty levels would become confounded between subject knowledge content and thinking skill being tapped.

Rich Naccarato also described a newer version of the Item Bank survey which will be going out in the near future.

A major portion of the remaining session time was consumed by the description of various taxonomies including Bloom's, Northwest Evaluation Association's, and the Center for the Study of Evaluation. Specific subparts of these taxonomies were compared and contrasted. The main message(s) behind this comparison was:

to point out similarities between taxonomies

to show that **none** of the taxonomies clearly depicted all higher-order thinking skills levels.

Samples were given of "higher-order" items that exist within a separate bank of 200-300 items belonging to the NWEA, and there was some discussion of how these would be classified within an item bank. Another product of the NWEA was discussed, the upcoming Science Item Bank, which will have taxonomical classifications that will be more amenable to selecting HOTS items within the science subject area. Participants were later encouraged to distribute the Item Bank survey form to anyone whom they knew currently was using a bank, particularly if the bank included HOTS items. Numerous handouts were presented, including an item bank survey form which is found in Naccarato (1987) "A Guide to Item Banking in Education, 3rd Edition".

ITEM BANK DATA SHEET

NACCARATO/MILLER
HANDOUT

Northwest Regional Educational Laboratory

Please type or write legibly. This questionnare will be photocopied.

Item Bank Title (if any) _____

For information about the bank contact:

Name _____
Title _____
Organization _____
Street _____
City _____ State _____ Zip _____
Phone: Area code _____ Number _____

Please indicate the characteristics of this item bank by placing an "X" in the appropriate boxes.

1. Which of the following are available through this bank? (check all that apply)

- ☐ test items
- ☐ classification of items by content
- ☐ general objectives or topic statements
- ☐ item specifications, detailed content descriptions, etc.
- ☐ suggested instructional activities
- ☐ cross references between objectives and appropriate instructional materials
- ☐ content review or other validity information
- ☐ reliability estimates
- ☐ p-values
- ☐ IRT (latent trait) calibrations
- ☐ other item analysis data
- ☐ technical reports

2. What is the source of the bank's objectives and items? (check all that apply)

- ☐ developed by teachers
- ☐ developed by state or local central office staff
- ☐ developed by test development personnel within your organization
- ☐ developed by an outside organization
- ☐ collected from other sources

3. What reviews or studies (if any) were performed for the items in the bank? (check all that apply)

- ☐ review to verify appropriate content
- ☐ content review to match items to objectives
- ☐ content review to establish appropriate grade levels or age levels
- ☐ editing for clarity
- ☐ editing based on reviews by technical personnel
- ☐ editing based on technical data
- ☐ review for sex bias
- ☐ review for cultural and ethnic bias
- ☐ informal pilot testing (informal selection of subjects, small numbers of subjects, non-rigorous analysis of results)
- ☐ formal pilot testing (rigorous sampling and anlysis of results, large numbers of subjects, standardized administration)

4. Which of these statements are correct for this bank?

YES	CAN BE ARRANGED	NO	
☐	☐	☐	all or most of the bank is available for sale
☐	☐	☐	all or most of the bank is available free (or for cost of reproduction)
☐	☐	☐	pre-developed tests are constructed from the item bank
☐	☐	☐	tests are constructed based on objectives in the item bank selected by the user
☐	☐	☐	items are used to construct tests based on objectives developed by the user

5. What additional services related to the bank can be provided? (check all that apply)

YES	CAN BE ARRANGED	NO	
☐	☐	☐	printing of test materials
☐	☐	☐	test scoring services
☐	☐	☐	development of individual student profiles
☐	☐	☐	development of class and school profiles
☐	☐	☐	training on test administration procedures
☐	☐	☐	training on writing or selecting objectives and item specifications
☐	☐	☐	training on writing test items
☐	☐	☐	assistance in interpreting test data

other _____
_____
_____
_____

6

Please place "Xs" in the appropriate spaces to indicate the subject areas and grade levels covered by this item bank. Also, please use the columns on the left to indicate the approximate numbers of items and objectives available in each content area. If the appropriate content area is not listed, please write it in the space provided.

A. When specifying grade levels, place an "X" in any range where at least one grade of the range is covered. For example, if your items cover grades 2 through 4, place "Xs" in both the K–2 and 3–5 columns.

B. We do not want to limit the content of the catalog to basic skills item banks. If your item bank deals with other content areas, please be sure to include them.

C. We realize that "Higher Order Thinking Skills" items may overlap other content areas within each subject.

	CONTENT AREA	APPROXIMATE NUMBER OF OBJECTIVES	APPROXIMATE NUMBER OF ITEMS	GRADE LEVEL (AGE) Pre School (0-4)	K-2 (5-7)	3-5 (8-10)	6-8 (11-13)	9-12 (14-17)	College	Adult
READING	Phonetic Analysis _____									
	Structural Analysis _____									
	Vocabulary _____									
	Comprehension _____									
	Reading Readiness _____									
	Reference (Study) Skills _____									
	Higher Order Thinking Skills _____									
	_____									
MATHEMATICS	Computational skills _____									
	Concepts _____									
	Problem Solving (Application) _____									
	Geometry _____									
	Calculator Math _____									
	Higher Order Thinking Skills _____									
	_____									
LANGUAGE ARTS	Grammar _____									
	Usage _____									
	Mechanics _____									
	Foreign Language (specify: _____)									
	Composition _____									
	Higher Order Thinking Skills _____									
	_____									
SCIENCE	_____									
	_____									
	Higher Order Thinking Skills _____									
OTHER	_____									
	_____									
	Higher Order Thinking Skills _____									

7. Which of the following can be used to retrieve items? (Check all that apply)

- ☐ objective to be tested
- ☐ item difficulty
- ☐ item type (e.g., multiple-choice)
- ☐ cognitive level (e.g., recall, inference)
- ☐ key words
- ☐ other_____
- ☐ none

8. If your items are retrievable by cognitive level, please indicate from where your cognitive categories were derived (If you would like to attach a listing of the taxonomy you use, we would very much appreciate it.)

- ☐ Bloom's Taxonomy
- ☐ other well-known cognitive taxonomy, please list _____
- ☐ modified an existing taxonomy, please list _____
- ☐ developed our own cognitive taxonomy
- ☐ other _____

9. In what form are the following available? (Check all that apply)

	HARD COPY	MICRO DISK	MAINFRAME DISK OR TAPE	NOT AVAILABLE
Item texts	☐	☐	☐	☐
Item graphics	☐	☐	☐	☐
Item statistics	☐	☐	☐	☐

10. If computer software is used to support item banking, please answer the following questions. (In each case, check all responses that apply.)

a. What function(s) does this software perform?

- ☐ item management
- ☐ test development
- ☐ scoring
- ☐ reporting
- ☐ student recordkeeping
- ☐ cross-reference to materials
- ☐ on-line test administration
- ☐ other _____

b. From where did you obtain this software?

- ☐ adapted from existing spreadsheets, databases, etc.
- ☐ If you checked the above box, please list the programs you are using
- ☐ specially designed for our system by
- ☐ purchased from test or research organization

c. If your software handles item management, indicate its capabilities below:

- ☐ stores items
- ☐ full-screen editing
- ☐ can add/delete/change items
- ☐ can add/delete/change item classifications
- ☐ will handle user history
- ☐ will handle test analysis
- ☐ other_____

d. If your software handles test development, indicate its capabilities below:

- ☐ on-line generation of tests
- ☐ can use various criteria for selecting items
- ☐ can add/delete/change items that are computer selected
- ☐ automatically stores answer keys
- ☐ prints tests
- ☐ has special print features
- ☐ will print multiple forms
- ☐ other _____

e. If your software handles scoring, please indicate its capabilities below:

- ☐ mark sense reader
- ☐ subtest and total scores
- ☐ objective mastery
- ☐ item statistics
- ☐ test statistics
- ☐ other _____

f. If your software handles reporting, please indicate its capabilities below

- ☐ subtest profiles--student
- ☐ subtest profiles--class
- ☐ summary reports
- ☐ cross reference to materials/methods
- ☐ other_____

g. If your software handles student recordkeeping, please indicate its capabilities below:

- ☐ by objective mastered
- ☐ total scores over time
- ☐ gradebook or series of scores
- ☐ other _____

h. If your software has graphics capabilities, please indicate these below:

- ☐ can generate most graphics
- ☐ can attach external graphics to item text
- ☐ can generate graphics and merge them in with item text
- ☐ can produce the entire test

i. What type of computer do you use? (If micro, please indicate XT, AT, 286, etc.)

_____

j. What is the name of your item banking/test scoring software package?

_____

k. Is your software:?

- ☐ available on exchange agreement
- ☐ available for purchase from vendor
- ☐ available for purchase from you
- ☐ not available for others

l. What is the approximate cost of your testing software?

_____

m. Technical assistance in running the software is

- ☐ available from us
- ☐ available through vendor
- ☐ not available

11. Please provide any descriptive comments or explanatory information about your item bank and/or testing system in this space or on a separate sheet of paper. We are especially interested in items measuring higher order thinking skills, and in computerized systems

From: Millman & Arter, 1984

Questions to be Answered in Designing Item Banking Systems

I. ITEMS

A. Acquisition and Development

1. Develop/use your own item collection or use collections of others?

   a. If develop your own item collection, what development procedures will be followed?

   b. If use collections of others, will the items be leased or purchased, and is the classification scheme sufficiently documented and the item format specifications sufficiently compatible for easy transfer and use?

2. What types of "items" will be permitted?

   a. Will open-ended (constructed response) items, opinion questions, instructional objectives, or descriptions of performance tasks be included in the bank?

   b. Will all the items be made to fit a common format (e.g., all multiple-choice with options a, b, c, and d ?

   c. Must the items be calibrated, validated, or otherwise carry additional information?

3. What will be the size of the item collection?

   a. How many items per objective/subtopic (collection depth)?

   b. How many different topics (collection breadth)?

4. What review, tryout and editing procedures will be used?

   a. Who will perform the review/editing?

   b. Will there be a field tryout, and if so, what statistics will be gathered, and what criteria will be used for inclusion into the bank?

B. Classification

1. How will the subject matter classifications be conducted?

   a. Will the classification by subject matter use fixed categories, keywords, or some combination of the two?

b. Who will be responsible for preparing the taxonomy?

c. How detailed will the taxonomy be? Will it be hierarchically or nonhierarchically aranged?

d. Who will assign classification indices to each item, and how will this assignment be verified?

2. What other assigned information about the items will be stored in the item bank? (See the attached list for potential attributes.)

3. What measured information about the items will be stored in the bank? (See the Appendix B list for potential measures.) How will the item measures be calculated?*

C. Management

1. Will provision be made for updating the classification scheme and items? If so:

a. Who will be permitted to make additions, deletions, and revisions?

b. What review procedures will be followed?

c. How will the changes be disseminated?

d. How will duplicate (or near duplicate) items be detected and eliminated?

e. When will a revision of an item be trivial enough that item statistics from a previous version can be aggregated with revisions from the current version?

f. Will item statistics be stored from each use, last use, or aggregated across uses?

2. How will items that require pictures, graphs, special characters, or other types of enhanced printing be handled?

3. How will items that must accompany other items, such as a series of questions about the same reading passage, be handled?

---

*This question is the subject of considerable controversy and discussion in the technical measurement literature. For example, to obtain a latent trait difficulty parameter, concern has been expressed about sample size, calibration procedure (Rasch, 3-parameter), linking models (major axis, least squares, maximum likelihood). and number of items common to the equating forms.

II. TESTS

A. Assembly

1. Must the test constructor specify the specific items to appear on the test or will the items be selected by the computer?

2. If the items are selected by the computer:

   a. How will one item out of several that matches the search specification be selected (randomly, time since last usage, frequency of previous use)?

   b. What happens if no item meets the search specifications?

   c. Will a test constructor have the option to reject a selected item, and if so, what will be the mechanism for doing so?

   d. What precautions will be taken to insure that examiners who are tested more than once do not receive the same items?

3. What item or test parameters can be specified for test assembly (item format restrictions, limits on difficulty levels, expected score distribution, expected test reliability, etc.)?

4. What assembly procedures will be available (options to multiple-choice items placed in random order, the test items placed in random order, different items on each test)?

5. Will the system print tests or just specify which items to use? If the former, how will the tests be printed or duplicated and where will the answers be displayed?

B. Administration, Scoring and Reporting

1. Will the system be capable of on-line test administration? If so:

   a. How will access be managed?

   b. Will test administration be adaptive, and if so, using what procedure?

2. Will the system provide for test scoring? If so:

   a. What scoring formula will be used (rights only, correction for guessing, partial credit for some answers, weighting by discrimination values)?

b. How will constructed responses be evaluated (off-line by the instructor, on-line/off-line by examiners comparing their answers to a key, on-line by computer with/without employing a spelling algorithm)?

3. Will the system provide for test reporting? If so:

a. What records will be kept (the tests themselves, individual student item responses, individual student test scores, school or other group scores) and for how long? Will new scores for individuals and groups supplement or replace old scores.

b. What reporting options (content/format) will be available?

c. To whom will the reports be sent?

C. Evaluation

1. Will reliability and validity data be collected? If so, what data will be collected by whom, and how will they be used?

2. Will norms be made available and, if so, based on what norm-referenced measures?

III. SYSTEM

A. Acquisition and Development

1. Who will be responsible for acquisition/development, given what resources, and operating under what constraints?

2. Will the system be made transportable to others? What levels and what degree of documentation will be available?

B. Software/Hardware Features

1. What aspects of the system will be computer assisted?

a. Where will the items be stored (computer, paper, card file)?

b. Will requests be filled using a batch, on-line, or manual mode?

2. Will items be stored as on large collection or will separate files be maintained for each user?

3. How will the item banking system be constructed (from scratch; by piecing together word processing, data-base management, and other general purpose programs; by adopting existing item banking systems)?

4. What specific equipment will be needed (for storage, retrieval, interactions with the system, etc.)?

5. How user and maintenance friendly will the equipment and support programs be?

6. Who will be responsible for equipment maintenance?

C. Monitoring and Training

1. What system features will be monitored (number of items per classification category, usage by user group, number of revisions until a user is satisfied, distribution of test lengths or other test characteristics, etc.)

2. Who will monitor the system, train users, and give support (initially, ongoing)?

3. How will information about changes in system procedures be disseminated?

D. Access and Security

1. Who will have access to the items and other information in the bank (authors/owners, teachers, students)? Who can request tests?

2. Will users have direct access to the system or must they go through an intermediary?

3. What procedures will be followed to secure the contents of the item bank (if they are to be secure)?

4. Where will the contents of the item bank be housed (centrally or will each user also have a copy)?

5. Who will have access to score reports?

IV. USE AND ACCEPTANCE

A. General

1. Who decides to what uses the item bank will be put? And will these uses be the ones that the test users need and want?

2. Who will develop the tests and who will be allowed to use the system? Will these people be acceptable to the examinees and recipients of the test information?

3. Will the system be able to handle the expected demand for use?

4. Will the output of the system likely to be used and used as intended?

5. How will user acceptance and item bank credibility be enhanced?

B. Instructional Improvement. If this is an intended use:

1. Will the item bank be part of a larger instructional/decision-making system?

2. Which textbooks, curriculum guidelines, and other materials, if any, will be keyed to the bank's items? Who will make that decision and how will the assignments be validated?

3. Will items be available for drill and practice as well as for testing?

4. Will information be available to users that will assist in the diagnosis of educational needs?

C. Adaptive Testing. If this is an option:

1. How will the scheduling of the test administrations take place?

2. How will the items be selected to insure testing efficiency yet maintain content representation and avoid duplication between successive test administrations?

3. What criteria will be used to terminate testing?

4. What scoring procedures will be followed?

D. Certification of Competence. If this is an intended use:

1. Will the item bank contain measures that cover all the important component skills of the competence being assessed?

2. How many attempts at passing the test will be allowed; when? How will these attempts be monitored?

E. Program/Curriculum Evaluation. If this is an intended use:

1. Will it be possible to implement the system so as to provide reliable measures of student achievement in a large number of specific performance areas?

2. Will the item bank contain measures that cover all the important stated objectives of the curriculum? That go beyond the stated objectives of the curriculum?

3. Will the item bank yield commensurable data that permit valid comparisions over time?

F. Testing and Reporting Requirements Imposed by External Agencies. If meeting these requirements is an intended use:

1. Will the system be able to handle requirements for program evaluation (e.g., Chapter 1), student selection into specially funded programs, assessing educational needs, and reporting?

2. Will the system be able to accomodate minor modifications in the testing and reporting requirements?

V. COSTS

A. Cost Feasibility

1. What are the (fixed, variable) costs (financial, time, space, equipment and supplies) to create and support the system?

2. Are these costs affordable?

B. Cost Comparisions

1. How do the item banking system costs compare to the present or other testing systems that achieve the same goals?

2. Do any expanded capabilities justify the extra cost? Are any restricted capabilities balanced by cost savings?

ITEM RETRIEVAL CHARACTERISTICS

1. Item Difficulty (classical or otherwise)
2. Question Format (M–C, T–F, etc.)
3. Reference to Curriculum
4. Readibility Level
5. Item Identification Number
6. Directions Required (Oral, etc.)
7. Cognitive Level (Recall, Inference, Application, Problem–Solving)

OTHER POSSIBLE CHARACTERISTICS

8. Source of the Item
9. Cross–referencing to Other Goals
10. Date of last use
11. I.D. of Item Writer/Editor
12. Security Level Code
13. Recommended Grade Level
14. Content Key Words

RESOURCES FOR INFORMATION ON ITEM BANKING

Arter, J. A. & Estes, G. D. (Nov., 1985). **Item banking for local test development: Practitioner's handbook.** Published by the Northwest Regional Educational Laboratory, Portland, OR.

Deck, D., Nickel, P. & Estes, G. (Nov.,1985). **Reviews of microcomputer item banking software.** Published by the Northwest Regional Educational Laboratory, Portland, OR.

Educational Testing Service (1984). **Item pools.** A publication of the ETS Test Collection, Educational Testing Service, Princeton, NJ.

Estes, G. D. (Ed.) (Nov., 1985). **Examples of item banks to support local test development: Two case studies with reactions.** Published by the Northwest Regional Educational Laboratory, Portland, OR.

Estes, G.D. & Arter, J.A. (1984). **A guide to item banking in education (second edition):Item bank data sheets.** Published by the Northwest Regional Educational Laboratory, Portland, OR.

Millman, J. & Arter, J. A. (Winter, 1984). Issues in item banking. **Journal of Educational Measurement,** 21, pp. 315-330.

Cross References for three HOTS SCHEMES for Cognition

o.	No. CSE – UCLA	No.	NWEA Uses Tricounty Course Goal Knowledge/Process Classifications	No.	Bloom's Taxonomy
1.	Mnemonic strategies				
2.	Cognitive strategies				
3.	Cognitive styles				
4.	Communication skills	P71–79	Vocalizing, moving, gesturing, touching, speaking, writing, media, dramatizing, etc.		
5.	Translation & Interpretation of Ideas	P–37	Summarizing, Abstracting	210 220	Comprehension–Translation Extrapolation
6.	Comparing	P–33	Classifying, categorizing, grouping, selecting according to criteria	1.20	Knowing ways & means to deal with specifics
7.	Sequencing	P–34	Ordering, sequencing	1.20 4.20	See above Analysis of organizational principles
8.	Classifying	P–33	Classifying, categorizing, grouping, selecting according to criteria	1.30	Knowledge of universals and abstractions in a field
9.	Categorizing	P–33	Same as above	1.30	See above
10.	Inventing–ideating	P–67	Creating on basis of knowledge and process	5.00	Synthesis
11.	Planning	P–34–36	P.61–63 Ordering, sequencing, arranging, transform, estimate, summarize, decide, solve, adapt	5.20	Production of a plan
12.	Problem redefinition	P–54	Revising Hypothesis	5.30	Derivation of a set of abstract relations
13.	Transfer of approach to new problems	P–65	Restructuring behavior, adapting, modifying	3.00	Application
14.	Creating Hypothesis	P–52	Formulating hypothesis	5.30	Derivation of a set of abstract relations
15.	Inferring from data	P–42	Inferring, interpolating, extrapolating	5.10	Production of unique communications
16.	Deductive Reasoning	P–48	Generalizing	6.00	Evaluation using internal/ external evidence
17.	Evaluating according to logical criteria	P–22	Evaluating logical consistency and accuracy	4.00 6.00	Analysis Evaluation
18.	Specifying judgemental criteria	P–47	Testing against standards or criteria	3.00 6.00	Application Evaluation
19.	Sensitivity to missing, irrelevant or misleading information	P–22	Evaluating authoritativeness, logical consistency, relevance, and adequacy	6.00	Evalulation in terms of internal/ external evidence
20.	Discriminating statements of value and fact	P–47	Testing against standards, criteria classifying, selecting according to criteria	6.00	Same as above
21.	Creating situations to provide data for evaluation	P–53	Testing hypothesis	5.20	Production of a plan or proposed set of operations
22.	Building theories or models	P–51	Theorizing, Predicting	5.30	Derivation of a set of abstract relations
23.	Valuing (internalizing decisions/plans)	P–64	Restructuring values (adapting, modifying)		

NWEA Reading Bank Index

PASSAGES

- 5.2 Reading
  - 5.2.1 Readiness
  - 5.2.2 Knowledge of Conventions
  - 5.2.3 Word Recognition, Word Meaning
    - 5.2.3.3 Structural Analysis
    - 5.2.3.4 Context Clues
    - 5.2.3.5 Homonyms, Synonyms, Antonyms, Acronyms, Malapropisms
  - 5.2.4 Literal Reading Comprehension
    - 5.2.4.1 Details
    - 5.2.4.2 Main Ideas
    - 5.2.4.3 Sequence
    - 5.2.4.4 Classification
    - 5.2.4.5 Directions
  - 5.2.5 Interpretive Comprehension
    - 5.2.5.1 Inference
    - 5.2.5.2 Analysis
    - 5.2.5.3 Association, Cause-Effect
    - 5.2.5.4 Comparison, Contrast, Analogy
    - 5.2.5.5 Summary, Synthesis
    - 5.2.5.6 Generalization
    - 5.2.5.7 Prediction
  - 5.2.6 Critical/Evaluative Comprehension
    - 5.2.6.1 Fact-Opinion
    - 5.2.6.2 Persuasion
    - 5.2.6.3 Internal/External Validation
    - 5.2.6.4 Assumptions, Author's Attitudes
  - 5.2.7 Applications

3452Psa
3/5/86

-Knowledge Categories-

- G1 Principles and Laws
- G2 Simple Generalizations
- K1 Conventions: Names and Nomenclature
- K2 Conventions: Symbols, Rules, Standardized Processes, Definitions
- K3 Properties, Parts, Characteristics, Features, Elements, Dimensions
- K4 Trends and Sequences
- K5 Similarities and Differences, Discriminations, Classifications
- K6 Contexts, Locations, and Orientations
- K7 Operations, Methods of Dealing with, Functions
- K8 Cause and Effect Relationships (Costs and Benefits)
- K9 Criteria or Standards
- K10 Non Cause-Effect Relationships

-Inquiry-Problem Solving Processes-

P1 Input — Acquiring Information

- P11 Viewing
- P12 Hearing
- P13 Feeling (tactile)
- P14 Smelling
- P15 Tasting
- P16 Using sense extenders
- P17 Using internal sensors of emotion

P2 Input Verification — Insuring Validity and Adequacy

- P21 Evaluating authoritativeness of sources
- P22 Evaluating logical consistency and accuracy
- P23 Evaluating relevance to desired learning purposes
- P24 Evaluating adequacy for acting or deciding (comprehensiveness and depth)

P3 Preprocessing — Organizing Information

- P31 Labeling, naming, numbering, coding
- P32 Recording, listing
- P33 Classifying, categorizing, grouping, selecting, according to criteria
- P34 Ordering, sequencing
- P35 Manipulating, arranging, transforming, computing
- P36 Estimating
- P37 Summarizing, abstracting

P4 Processing I — Interpreting Information (drawing meaning from data)

- P41 Decoding verbal and non-verbal symbols (reading and literal translating)
- P42 Inferring, interpolating, extrapolating
- P43 Analyzing
- P44 Associating, relating, equating
- P45 Comparing, contrasting, discriminating
- P46 Synthesizing
- P47 Testing against standards or criteria
- P48 Generalizing

P5 Processing II — Using Information to Produce New Information

- P51 Theorizing, predicting
- P52 Formulating hypotheses
- P53 Testing hypotheses
- P54 Revising hypotheses

P6 Output I — Acting on the Basis of Information

- P61 Reacting
- P62 Making decision
- P63 Solving problems
- P64 Restructuring values (adapting, modifying)
- P65 Restructuring behavior (adapting, modifying)
- P66 Encoding verbal and nonverbal symbols prior to communication
- P67 Creating on the basis of knowledge and process

P7 Output II — Communicating Information

- P71 Vocalizing (nonverbal)
- P72 Gesturing, moving
- P73 Touching
- P74 Speaking
- P75 Writing
- P76 Using art media (painting, drawing, sculpting, constructing, etc.)
- P77 Dramatizing
- P78 Singing, playing instruments
- P79 Dancing

Alternative Formats For Assessing Higher Order Thinking Skills

Edys S. Quellmalz

The purpose of Dr. Quellmalz's presentation was to provide the audience with a variety of alternative formats for assessing HOTS. She began by describing categories of HOTS skills and processes that she has found useful when attempting to assist districts to examine HOTS. This list is provided on page 1 of her handouts.

- **Analyze** -- Identify components of things; what something is made up of.
- **Compare** -- Compare properties, events, features of two or more things.
- **Infer/Interpret** -- Draw conclusions.
- **Evaluate** -- Apply criteria to make and defend solutions or judgements.

Dr. Quellmalz finds that these thinking processes cut across all content areas and that the terminology used by philosophers to describe "critical thinking", psychologists to describe "problem solving", etc. both really describe these same basic set of processes. These points are illustrated by pages 2 and 3 of her handouts. Figure 1 shows how terminology across disciplines can be aligned; Figure 3 shows how the HOTS processes described above apply to various content areas.

Before developing assessment questions that measure these processes, one must make several decisions.

1. Will skills be measured in isolation or will the integrated use of skills be emphasized? This point refers to the issue of reductionism discussed in the keynote address.
2. How will metacognitive skills be handled? Metacognitive skills are listed on page 1 of Dr. Quellmalz's handout.
3. Will the content of the questions deal with significant issues, recurrent issues or novel situations with which the student must deal?
4. How will prior knowledge be handled? Will you deal with situations that should be within the personal experience of students; information that students have been taught at school; or will all information be provided in the test itself?
5. How will questions be presented? This includes presentation mode such as written, visual, or demonstration; the amount of information in the question that must be processed by the student; whether information will be provided on the criteria by which a response will be evaluated; and response format such as written, oral, performance, group, individual.
6. Will the test be timed?
7. How will the assessment be tied into instruction. Dr. Quellmalz provided an example (see page 4 of her handout) of how a set of criteria and a listing of what needs to be included when one uses a HOTS process can be used to both guide instruction and evaluate student responses.

Dr. Quellmalz then proceeded to show a variety of questions that she has helped developed that attempt to assess the various HOTS skills and processes in a number of content areas. The following represents a few of those presented:

1. **Essay.** An example would be "Compare the Russian and U.S. versions of the Russian revolution." For this type of question, Dr. Quellmalz includes extra information for the student -- passages from a Russian and U.S. textbook, what should be covered (such as "what would they both agree on, disagree on and why"), and some information about how it will be scored (such as "be sure to support your arguments with examples from the text" or "be sure to explain your position, describe your criteria for deciding on a position, and provide examples which support your opinion").

2. **Partial Essay.** In this question type, part of the essay is already written. The student is asked to complete it by adding supporting examples, etc. This format might be useful for students just learning how to develop arguments.

3. **Combination multiple-choice, short answer and essay.** Combining the question types covering a presented situation might allow for a more complete measures. Multiple-choice questions could cover finding information and other HOTS skills measurable in multiple-choice format short answer questions and essays make students explain their reasoning. Also increasing the number of questions asked about a situation enables one to provide a more complex situation.

In preparation for the next session that Dr. Quellmalz presented jointly with Dr. Rick Stiggins, she asked participants to classify the questions on pages 5 and 6 of her handout as to what level of thinking skill each one tapped.

FIGURE 2

HIGHER ORDER THINKING

Students engage in proposeful, extended lines of thought where they:

- Identify the task of problem type
- Define and clarify essential elements and terms
- Judge and connect relevant information
- Evaluate and adequacy of information and procedures for drawing conclusions and/or solving problems

In addition, students will become self-conscious about their thinking and develop their self-monitoring problem solving strategies.

COMMONLY SPECIFIED
HIGHER ORDER REASONING PROCESSES

COGNITIVE

Analyze

Compare

Infer/Interpret

Evaluate

METACOGNITIVE

Plan

Monitor

Review/Revise

FIGURE 1

RELATIONSHIP AMONG REASONING SKILLS PROPOSED IN PSYCHOLOGY AND PHILOSOPHY

CRITICAL THINKING SKILLS (Philosophy)	PROBLEM SOLVING STRATEGIES (Psychology)	PROBABLE DOMINANT COGNITIVE PROCESSES (Psychology)
Clarification - Identify or formulate a question - Analyze major components - Define important terms	- Identify the problem - Identify essential elements and terms	Analogical analysis comparison
Judge credibility of support, the source, observations	Identify appropriate information, content and procedural schemata	Analogical analysis comparison evaluate components
Inference - deduction - induction - value judgments - fallacies	Connect and use information to solve the problem	Inferential infer/interpret relationships among components
Use criteria to judge adequacy of solutions	Evaluate success of the solution	Evaluative effectiveness of specific and general strategies

Figure 3. Examples of Higher-Order Reasoning Skills in Three Subject Domains

		Science	Social Science	Literature
1.	Analyze	Identify the components of process and the features of animate and inanimate objects	Analyze components or elements of an event	Identify components of literary, expository, and persuasive discourse
2.	Compare	Compare the properties of objects or events	Compare causes and effects of separate events; compare social, political, economic, cultural, and geographic features	Compare meanings, themes, plots, characters, settings, and reasons
3.	Infer	Draw conclusions; make predictions; pose hypotheses, tests, and explanations	Predict, hypothesize, and conclude	Infer characters' motivation; infer cause and effect
4.	Evaluate	Evaluate soundness and significance of findings	Evaluate credibility of arguments, decisions, and reports; evaluate significance	Evaluate form, believability, significance, completeness, and clarity

HIGHER ORDER THINKING (H.O.T.) PROJECT

THINKING STRATEGIES: COMPARE

1. Identify the objects, events, ideas to be compared
2. Identify relevant examples/attributes that are similar or different to compare
3. Identify/list categories of attributes that are similar or different
4. Explain the reasons or function of the similarities and/or differences (so what?)

EVALUATIVE CRITERIA FOR STUDENT RESPONSES

1. Names objects/events/ideas to be compared	1 no	2 implies	3 names	4 defines
2. Identifies categories for comparison - are they distinctive? - is there a sufficient number?	1 no	2 few	3 some	4 sufficient
3. Appropriately identifies and explains how the objects/events/ideas compare and/or differ on attributes within each category - are the comparisons accurate? - are they balanced i.e. describe differences? - how thorough are they?	1 little or none	2 few	3 some	4 most all
4. Presents comparisons in a logical sequence	1 confused	2 some off	3 mostly	4 almost all
5. Explains the significance/function of the similarities and/or differences - how clearly? - how appropriately? - how extensively?	1 not at all	2 somewhat	3 some	4 well

Final Step: A Progress Check

In the space provided next to each exercise, enter the letter that represents the thinking skill category reflected in the item (See Appendix B for answers):

R = Recall
A = Analysis
C = Comparison
I = Inference
E = Evaluation

____ 1. What are 3 functions of the liver?

____ 2. Let's brainstorm what would happen if the sun did not come up tomorrow.

____ 3. Define the word mitasis.

____ 4. Which of the following menus is the best? Why?

____ 5. Which menu provides more complete protein?

____ 6. Should the use of computers be abolished in the classroom? Why or why not?

____ 7. Who is the author of Where the Sidewalk Ends?

____ 8. If we mix these chemicals together, what do you suppose will happen?

____ 9. Look at the chart showing the number of meals Americans have eaten away from home in the last three years. How have eating habits changed?

____ 10. What are three purposes of an unmanned space flight to Jupiter?

____ 11. What are the functions of our eyelashes?

____ 12. Which do you think will have greater impact on your life, the invention of the computer or our ability to travel in space? Why?

____ 13. If you were going outside and it was snowing quite hard, which of the following would you need from your closet?

____ a) Your umbrella
____ b) A light weight jacket
____ c) Your warm boots
____ d) Your sandals

____ 14. You hate rain, but know it is necessary. What are three purposes it serves?

____ 15. In the Northwest it rains and snows a lot. Which is more vital for the necessary supply of water for summer use?

____ 16. What are some jobs a migrant worker might perform in getting a crop of lettuce to market?

____ 17. Haiku is a form of _____.

____ 18. Look at the three paintings. Which makes the most use of vivid colors?

____ 19. Suppose we had not dropped the bombs on Hiroshima and Nagasaki, how else might we have defeated the Japanese?

____ 20. Which is a better snack for you, a fresh peach or a dish of frozen peach yogurt? Why?

Classroom Assessment of Higher Order Thinking Skills

Richard Stiggins and Edys S. Quellmalz

Rick Stiggins and Edys Quellmalz teamed up to provide practical strategies for teaching teachers to measure higher order thinking skills in the classroom on a day-to-day basis. After reviewing a wide variety of reasons why each teacher must be skilled in this kind of assessment, the presenters introduced the Assessment Planning Chart, an easy-to-use planning device designed to raise teachers' level of consciousness about thinking skills and their measurement. The chart was taken from "Measuring Thinking Skills in the Classroom" by R. Stiggins, E. Rudel and E. Quellmalz.

Participants went through a series of steps in learning how to use the chart. Essentially, the chart crosses five levels of thinking skills (recall, analysis, comparison, inference, and evaluation) with three types of assessment (oral questions, test items, and performance assessments) to create five-by-three, fifteen cell matrices. Within each cell, the user writes an assessment probe tapping the appropriate level of thinking.

As a first step in the process of learning to use the chart, participants reviewed a completed chart, studying each cell entry to discern how and why it reflected a particular level of thinking. In step two, they filled in empty cells in a partially completed chart. In step three, they used a variety of prespecified question forms and formats to complete a chart on a topic of special interest to them. And finally, using a variety of other instructional aids, each participant prepared a complete chart on their own.

As a conclusion to the session, the participants brainstormed various ways to integrate this kind of assessment into actual classroom contexts. The presenters initiated the list of ideas based on their experience, and the participating teachers and administrators added to the extensive list of suggestions.

Table 1
SUMMARY OF THINKING SKILLS

Level	Definition	Relation to Bloom Taxonomy
Recall	Most tasks require that students recognize or remember key facts, definitions, concepts, rules, and principles. Recall questions require students to repeat verbatim or to paraphrase given information. To recall information, students need most often to rehearse or practice it, and then to associate it with other, related concepts. The Bloom taxonomy levels of knowledge and comprehension are subsumed here, since verbatim repetition and translation into the student's own words represent acceptable evidence of learning and understanding.	Recall Comprehension
Analysis	In this operation, students divide a whole into component elements. Generally the part/whole relationships and the cause/effect relationships that characterize knowledge within subject domains are essential components of more complex tasks. The components can be the distinctive characteristics of objects or ideas, or the basic actions of procedures or events. This definition of analysis is the same as that in the Bloom taxonomy.	Analysis
Comparison	These tasks require students to recognize or explain similarities and differences. Simple comparisons require attention to one or a few very obvious attributes or component processes, while complex comparisons require identification of and differentiation among many attributes or component actions. This category relates to some of the skills in the Bloom level of analysis. The separate comparison category emphasizes the distinct information processing required when students go beyond breaking the whole into parts in order to compare similarities and differences.	Analysis
Inference	Both deductive and inductive reasoning fall in this category. In deductive tasks, students are given a generalization and are required to recognize or explain the evidence that relates to it. Applications of rules and "if-then" relationships require inference. In inductive tasks, students are given the evidence or details and are required to come up with the generalization. Hypothesizing, predicting, concluding, and synthesizing all require students to relate and integrate information. Inductive and deductive reasoning relate to the Bloom levels of application and synthesis. Application of a rule is one kind of deductive reasoning; synthesis, putting parts together to form a generalization, occurs in both inductive and deductive reasoning.	Application Synthesis
Evaluation	These tasks require students to judge quality, credibility, worth, or practicality. Generally we expect students to use established criteria and explain how these criteria are or are not met. The criteria might be established rules of evidence, logic, or shared values. Bloom's levels of synthesis and evaluation are involved in this category. To evaluate, students must assemble and explain the interrelationship of evidence and reasons in support of their conclusion (synthesis). Explanation of criteria for reaching a conclusion is unique to evaluative reasoning.	Synthesis Evaluation

ASSESSMENT PLANNING CHART

GRADE LEVEL Junior High SUBJECT Social Studies TOPIC Electoral College

	ORAL	TEST	PERFORMANCE
RECALL	What is the electoral college?	As a member of the electoral college, you must vote: a. According to your own judgment b. As your constituency voted c. As the party tells you d. Only if you wish to do so.	Assume you're a U.S. senator. Propose a constitutional amendment that would make the popular vote the sole criterion for electing a president. Your amendment would do away with the electoral college. Prepare a speech to Congress defending your amendment. Be sure to— a. Analyze all elements of the issue, b. Compare elections with and without the college, c. Show how the voters are likely to react, and d. State and defend your values. (Or conduct a simulated debate on the Senate floor.)
ANALYSIS	How does the electoral college work?	Analyze the steps in the presidential election process, showing where the electoral college comes into play.	
COMPARISON	How do the social conditions that existed when the electoral college was formed differ from conditions now?	What is meant by the election theme "one person, one vote," and how does that relate to the electoral college?	
INFERENCE	If you were a presidential candidate elected by popular vote, could you still lose the election? How?	In which state is the electorate likely to oppose the use of the electoral college: a. California b. Illinois c. Etc.	
EVALUATION	Should the electoral college be abolished? Why or why not?	Which of the following is the best reason for maintaining the electoral college? a. Tradition. b. Fairness to large states. c. Etc.	

142

SIGGINS/QUESTIONS HANDOUT

ASSESSMENT PLANNING CHART

GRADE LEVEL High School SUBJECT Social Studies TOPIC Branches of Government

	ORAL	TEST	PERFORMANCE
RECALL	Who's in charge of the executive branch of government? The judicial branch? The legislature?		Set the class up as a democracy with three branches of government. Go through the actual process of passing a law. Determine if students know and can carry out each function. Strive to demonstrate the interactions among branches.
ANALYSIS		What can the executive branch do about an unfair law?	
COMPARISON	In America we have Congress. In Canada their legislature is called Parliament. How do they differ in structure? *OR* Compare the structure of the federal government with that of your state government.		
INFERENCE		You've decided motorcycles cause too many accidents. You'd like them banned from federally funded highways. To which government branch would you first appeal? a. Executive b. Judicial c. Legislative	
EVALUATION	Which branch of the government is the most important? Why?		

STIGGINS/QUELLMALZ
HANDOUT

ASSESSMENT PLANNING CHART

GRADE LEVEL _____ SUBJECT Generic Chart TOPIC _____

	ORAL	TEST	PERFORMANCE
RECALL	What is _____?	What is the best definition for the term _____? (a) (b) (c)	Give a speech/plan a debate in which you use what you know about _____ to judge its _____. Give reasons to explain your point of view.
ANALYSIS	How does _____ work?	What are the basic elements (ingredients) of _____?	
COMPARISON	Compare the _____ to _____.	What is the major difference between _____ and _____? (a) (b) (c)	
INFERENCE	What do you think would happen if _____? When _____ happens, do _____.	What is the main point of _____? Which of the following is a likely result of _____? (a) (b) (c)	
EVALUATION	Is this a satisfactory solution to this problem _____? How would you do it? (Follow up: why?)	Here is the issue: _____. Which side are you on and why?	

STIGGINS/QUELLMALZ
HANDOUT

ASSESSMENT PLANNING CHART

GRADE LEVEL _____ SUBJECT _____ TOPIC _____

	ORAL	TEST	PERFORMANCE
RECALL			
ANALYSIS			
COMPARISON			
INFERENCE			
EVALUATION			

145

STIGGINS/QUELLMALZ
HANDOUT

Study Step 4: Adding Variety to the Questions

Now lets move from the "formula" chart to the generation of charts with a greater variety of questions. The first key to expanding the range of questions you can pose is to focus on the trigger or action verb used to describe the problem to the student. Start with these and add some of your own if you can:

If you want to measure:	Use these key words in the exercise:		Illustration
Recall	define identify label list name repeat	what when who where	List the names of the main characters in the story.
Analysis	subdivide breakdown separate categorize sort	classify group	Break the story down into different parts.
Comparison	compare contrast differentiate distinguish		Compare themes of these two stories.
Inference	deduce predict infer speculate anticipate what if...	apply conclude	How might we make this character more believable?
Evaluation	evaluate judge assess appraise defend decide	argue recommend debate why critique	Evaluate this story. Is it well written? Why or why not?

The second key to expanding the range of questions you can pose is to plug these action words into a growing list of generic questions. Again, consider these and add some of your own if you can:

Recall

- Define the word _____.
- What is a _____?
- Label the following _____.
- Identify the _____ in this _____.
- Who did _____?

Analysis

- What are the basic elements (ingredients) in a _____?
- What is/are the functions of _____?
- Inventory the parts of _____.
- Categorize the _____ of _____.
- Sort the _____.
- Analyze the following _____.
- Classify _____.

Comparison

- Compare the _____ before and after.
- Contrast the _____ to the _____.
- Differentiate between _____ and _____.

Inference

- Hypothesize what will happen if _____.
- Predict what would be true if _____.
- Conclude what the result will be if _____.
- What if _____ had happened instead?

- What does this information suggest?
- Given this situation (problem) what should you do?
- What rule applies in this case?

Evaluation

- What do you believe about _____?
- Judge what would be the best way to solve the problem of _____. Why did you decide that?
- Evaluate whether you would _____ or _____ in this situation. Why?
- Decide if _____ was worth it. Explain.

Use these lists of action verbs and questions to generate a complete chart on another topic of relevance to you.

Assessment Strategies to High Order Thinking Skills

Catherine Ross

Catherine Ross opened her presentation by asking two key questions educators should consider when addressing critical thinking skills:

- What do we mean by evaluation?
- What behaviors do we seek to evaluate when we evaluate "higher level thinking skills"?

Experts in the field have many definitions of higher level thinking skills; Ross discussed several of them. She prefers the question posed by Dr. Arthur Costa: "How do students behave when they don't know?"

Ross then had the audience engage in an activity, which would also be useable in the classroom, where they "did not know" the answer, and had to use higher level thinking skills to determine it. The group then discussed the processes they had used to solve the problem. Ross suggested that these were the "behaviors" we seek to measure when we attempt to evaluate higher level thinking skills. She emphasized that classroom observation of students' growth in these skills can be as effective in evaluation as can the paper/pencil test. Teachers should look for such evidence as 1) increased perserverance when working on problems, 2) flexibility in trying different approaches to solving problems, 3) enjoyment of the process of working on problems that are not immediately solvable and 4) transference of problem-solving skills developed in one situation to another, dissimilar situation.

For the last portion of the presentation, Ross went through the eight sections of the Ross Test of Higher Cognitive Processes, giving sample items from each section:

Analogies
Missing Premises
Deductive Reasoning
Abstract Relations
Sequential Synthesis
Questioning Strategies
Relevant and Irrelevant Information
Analysis of Attributes

Appendix A

Conference Program

Assessing Higher Order Thinking Skills

Issues and Practices

October 1 and 2, 1987

Monarch Hotel
Clackamas, Oregon

sponsored by

Northwest Regional Educational Laboratory

and

ESD #112, Vancouver, Washington

Idaho Association of Secondary School Principals

Northwest Evaluation Association

Oregon Department of Education

Washington Office of the Superintendent

of Public Instruction

Welcome!

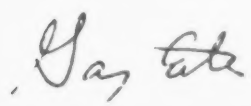
The Northwest Regional Educational Laboratory, in conjunction with five regional associations and departments of education, is pleased to welcome you to two days of presentations and workshops on assessing higher order thinking skills (HOTS). Good assessment is essential if we are going to make decisions about the instructional needs and progress of students. One goal of this conference is to assist people who are deciding how to assess this critical area in our schools to make more informed decisions and to produce higher quality assessments. Another goal of this conference is to bring together people interested in the assessment aspects of thinking skills. One outcome might be a cooperative effort in item banking.

The conference has been divided into two parts: Issues and Practices. The afternoon of the first day is designed to elicit perspectives about assessing higher order thinking skills. Issues such as how should HOTS be assessed, the quality of existing tests, how do we define HOTS and what will the future bring will be discussed. This part will set the stage for the second day's activities.

The second day emphasizes practices. There are three strands: daily, less formal classroom assessment; formal assessment procedures that may be used at the district level; and training on how to write test questions which assess HOTS.

The presenters and participants in the eighteen sessions will make each of the sessions a worthwhile exploration of the issues and practices for assessing higher order thinking skills.

Once again, welcome! We are looking forward to a productive conference.



Gary Estes
Director, Evaluation and Assessment
Northwest Regional Educational Laboratory

Conference Schedule at a Glance

Thursday, October 1

12:00 Registration

1:30 Welcome

1:45 Keynote Address: Stuart Rankin

2:45 BREAK

3:00 Panel Response: Kenneth Bumgarner, Stephen P. Norris, Edys S. Quellmalz

4:00 Discussion

4:45 Social

Friday, October 2

8:30 Refreshments

9:00 Writing Multiple-choice Critical Thinking Items (part 1): Stephen P. Norris

In-classroom, Informal Assessment of Students' Thinking Skills: Kenneth Bumgarner

How to Take Unlikely-Looking Textbook Material and Make it Testable for Critical Thinking: Connie Missimer

How to Select a Higher Order Thinking Skills Test: Judith A. Arter

10:15 BREAK

10:30 Writing Multiple-choice Critical Thinking Items (part 2): Stephen P. Norris

Diagnosing Thinking Deficiencies: Selma Wassermann

Assessing Creativity: Anita Halstead

Assessing HOTS in Hawaii: One State's Experience: Judith A. Arter and Lynde S. Paule

12:00 LUNCHEON

1:30 Alternative Item Formats for Assessing HOTS (part 1): Edys S. Quellmalz

Diagnosing Thinking Deficiencies: Selma Wassermann (repeat)

Test Publishers' Panel: Academic Therapy, Psychological Corporation, Riverside, and SRA

Item Banking and Higher Order Thinking Skills: Richard W. Naccarato and Ray K. Miller

2:45 BREAK

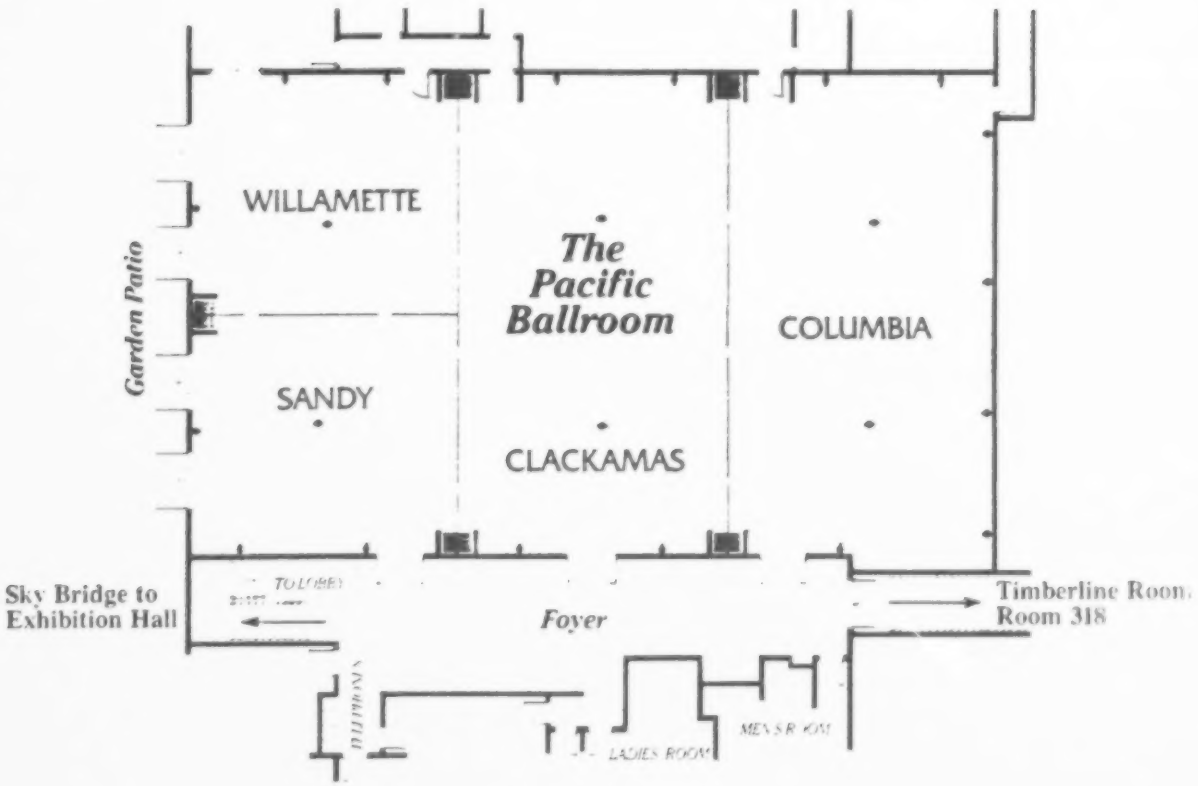
3:00 Alternative Item Formats for Assessing HOTS (part 2): Edys S. Quellmalz

Assessment Strategies for Higher Order Thinking Skills: Catherine Ross

Classroom Assessment of Higher Order Thinking Skills: Richard Stiggins

Designing a Testing Program: Wayne Neuberger

MONARCH HOTEL CONFERENCE ROOMS



DESCRIPTION OF SESSIONS

Thursday, October 1

1:45-3:00 **Keynote Address: Stuart Rankin**
Columbia Room
Rankin will identify the major issues that need attention in assessing student thinking and will suggest positions on those issues. He will consider: 1) content (What are the thinking skills?); 2) implications of assessment procedures and results for instruction; 3) assessment within or across disciplines; and 4) measurement questions.

3:15-4:30 **Panel Response and Discussion: Ken Bumgarner, Stephen P. Norris, and Edys Quellmalz. Moderator: Vida Taylor.**
Columbia Room

4:45 **Social Hour (no host)**
Sandy Room
You are invited to join workshop presenters and participants for hors d'oeuvres and a no host bar.

CONCURRENT WORKSHOP SESSIONS

Friday, October 2
9:30 - 10:15

Clackamas Room

Writing Multiple-Choice Critical Thinking Items (part 1): Stephen P. Norris.
This session will concentrate on writing multiple-choice items for assessing specific aspects of critical thinking such as induction, credibility of authorities and observation, and assumption identification. Explanations will be given for why certain sorts of item formats are not suitable for assessing these aspects of critical thinking. More suitable formats will be described and some practice time for writing items will be provided.

Willamette Room

In-classroom, Informal Assessment of Students' Thinking Skills: Kenneth Bumgarner.
This session will focus on the ways classroom teachers can assess whether their students' thinking skills are improving, using tools/techniques readily available to the classroom teacher, such as interviews, observation of student pair learning, use of thinking logs, etc. The assessment emphasis will be on "how students behave when they don't know the answer" -- as Art Costa puts it. Participants will be invited to share in-classroom techniques they use to get an idea of whether their students' thinking is improving.

Sandy Room

Sow's Ears Into Silk Purses - How to Take Unlikely-Looking Textbook Material and Make it Testable for Critical Thinking: Connie Missimer.
One of the biggest problems teachers face is the fact that many textbooks do not readily lend themselves to critical thinking. This workshop will be "hands-on," showing some techniques for pulling critical thinking out of unlikely places, then offering everyone the opportunity to try their hands at some additional examples.

Multnomah Room

How to Select a Test of Higher Order Thinking Skills: Judith A. Arter.
The Test Center staff at NWREL has developed a checklist designed to assist those interested in selecting a test of Higher Order Thinking Skills. The checklist includes considerations of content, reliability, validity and usability. These issues will be explored as they apply to currently available instruments and the issues discussed in the keynote address. A selection of HOTS assessment instruments will be available for examination by participants. This session would be appropriate for anyone currently trying to select a HOTS test for local use.

10:30 - 11:45

Clackamas Room

Writing Multiple-Choice Critical Thinking Items (part 2): Stephen P. Norris.
This session will describe an approach for obtaining information on the validity of multiple choice critical thinking items and tests. The approach depends upon asking samples of students to think aloud as they work through trial items. Decisions to modify, retain, or discard items are based upon comparisons between the quality of the students' reasoning and their choices of answers. Generally speaking, we want quality reasoning to be associated with keyed answers and poor quality reasoning to be associated with unkeyed answers. The presentation will illustrate how the approach was used in the design of a test of high school students' ability to assess the credibility of observations, but different grade levels and different aspects of critical thinking could have been chosen.

Willamette Room

Diagnosing Thinking Deficiencies in the Classroom: Selma Wassermann.
Most teachers implicitly understand the relationship between thinking and behavior. They see, in their day-to-day dealings with students that pupils may behave "thoughtlessly" or inappropriately. They observe, to their dismay, that some pupils persistently respond to their instructions by saying, "I don't get it" or "I don't understand what you mean." Teachers are quick to discern such lapses in thinking: "If you would think about that, you'd figure it out for yourself!" But urging a student to "think for himself" cannot correct those deficits in his higher order functioning that produce his dysfunctional behavior.

This workshop will examine eight dysfunctional classroom behavior patterns that give evidence of pupils' impaired higher order functioning. Based upon the work of Louis Raths, the workshop demonstrates the use of classroom observation instruments to assess pupils' "thinking-related" behavioral patterns and suggests how classroom materials and instructional strategies may work to reduce such dysfunctional classroom behavior.

Sandy Room

Assessing Creativity: Anita Halstead.
This session will examine numerous informal measurements to assess creative abilities in children and adults. The emphasis will be on practical applications in the classroom.

Multnomah Room

Assessing Higher Order Thinking Skills in Hawaii: One State's Experience: Judith A. Arter and Lynde S. Paule.
NWREL is currently under contract with the Hawaii Department of Education to develop multiple-choice questions that measure decision-making and problem solving skills. The presenter will discuss the procedures, pitfalls and special studies undertaken during the course of this development. Sharing of similar experiences by participants will be encouraged. This session would be appropriate for those currently considering developing their own HOTS test.

12:00
Columbia Room

LUNCHEON

1:30 - 2:45

Clackamas Room

Alternative Item Formats for Assessing HOTS (part 1): Edys S. Quellmalz
In this workshop participants will develop a variety of approaches for evaluating students' higher order thinking skills. Following an introduction to the rationale for the framework which specifies four broad categories of higher order skills: 1) analysis, 2) comparison, 3) inference, and 4) evaluation; participants will examine examples of alternative item formats. The workshop will then provide opportunities to develop and discuss assessment tasks designed to measure achievement in each of the four categories. The workshop will emphasize open-ended, constructed formats, and the use of higher order reasoning within on-going curricula.

Willamette Room

Diagnosing Thinking Deficiencies in the Classroom: Selma Wassermann.
Repeat of morning session.

Sandy Room

Test Publishers' Panel: Catherine Ross, Gale H. Roid, Vern Dahl, and Margaret Hill. Moderator: Ken Bumgarner.
Representatives from Academic Therapy, Psychological Corporation, Riverside and SRA will discuss what is currently available for assessing higher order thinking skills. Achievement tests as well as specific tests of higher order thinking (Watson-Glaser and the Ross Test of Higher Cognitive Processes) will be presented. A discussion will follow.

Multnomah Room

Higher Order Thinking Skills and Item Banking: Richard Naccarato and Ray K. Miller.
With the recent popularity of item banking at the state and local levels, one might question how difficult it would be to identify items that measure higher order thinking skills in the various subject areas. Is it be desirable to key test items on this dimension? Representatives from NWREL and NWEA will discuss, both formally and informally, their awareness of what seems to be happening nationally on this topic; some available item banks that are keyed to the HOTS dimensions; strategies for creating cross-referencing to these skills, and generating tests with HOTS dimensions in mind.

3:00 - 4:15

Clackamas Room

Alternative Item Formats for Assessing HOTS (part 2): Edys S. Quellmalz.
This is a continuation of the 1:30 session.

Willamette Room

Assessment Strategies for Higher Order Thinking Skills: Catherine Ross.
This session will look at assessing higher level thinking skills with sample items on: identifying analogous and metaphorical relationships; questioning strategies; attribute analysis; determining relevancy of data; and formulating/testing hypotheses.

Sandy Room

Classroom Assessment of Higher Order Thinking Skills: Richard Stiggins.
The session will be divided into two parts. The first part will focus on the results of a new study of the extent to which teachers in one school district actually measure Higher Order Thinking Skills in their classrooms on a day to day basis. The second part will be a discussion of strategies for teaching teachers to measure HOTS effectively and efficiently. In both cases, the focus will be on assessments devised by teachers or selected by teachers from available instructional materials.

Multnomah Room

Designing a Testing Program to Include Higher Order Thinking Skills: Wayne Neuberger.
This session will focus on the purposeful uses of tests. It will review the various purposes of testing, particularly as they apply to higher order thinking skills. It will also include procedures for reviewing current testing practices, and planning for modification of a testing program to address the purposes to be served. The session will focus on practical ways to review and modify testing programs.

Judith A. Arter, PhD
Northwest Regional Educational Laboratory
101 S.W. Main Street, Suite 500
Portland, Oregon 97204
Dr. Arter has been associated with NWREL for nine years. During that time she has been involved in several projects including provision of evaluation and testing technical assistance with the Chapter 1 Evaluation Technical Assistance Center, a national study of refugee adult ESL programs, and several test development projects for school districts, state departments of education and businesses. She is currently director of the Test Center at NWREL and has been co-author of two consumer guides designed to assist persons select assessment instruments in the areas of HOTS and School Climate. Dr. Arter has a bachelor's degree in mathematics from the University of California, San Diego; and a PhD in special education from the University of Illinois, Urbana-Champaign.

Kenneth Bumgarner, EdD
Division of Instructional Programs & Services
Old Capitol Building, FG-11
Olympia, WA 98504
Dr. Bumgarner is Director of Basic Education and Curriculum for the Office of the Superintendent of Public Instruction (Washington State SEA) and an adjunct faculty member at Seattle University and St. Martin's College (Olympia), where he teaches classes related to the improvement of students' thinking skills. He has had a primary role in leading the state's collaborative effort for improving student thinking in the classroom. He earned his EdD at Seattle University.

Vern Dahl
Riverside Publishing Company
P.O. Box 4273
Carmel, CA 93921
Mr. Dahl earned an MA in Educational Psychology from the University of Minnesota. He was a school counselor and teacher for eight years before entering the test publishing field. Mr. Dahl has worked for Harcourt Brace Jovanovich, Westinghouse, CTB/McGraw-Hill and recently was hired as an Executive Consultant for Riverside.

Anita Halstead
310 Union Avenue
Snohomish, WA 98290
Ms. Halstead has a bachelor's degree in Political Science from the University of Washington, Seattle, and a Master's degree in Creative Education from Seattle Pacific University. She has served as a research analyst for the CIA, a city councilwoman, college administrator and gifted education specialist. She originated *Creativity Network* in 1979, a unique monthly newsletter. In addition, she has written handbooks for managers and businesses, classroom curriculums, a children's workbook and has produced an audiocassette on creativity.

Margaret Hill
Science Research Associates, Inc.
155 North Wacker Drive
Chicago, IL 60606
Ms. Hill is the Manager of Test Development for SRA where she has worked in test publishing for 15 years. She has worked with the Survey of Basic Skills and with the SRA Item Bank. She earned a MS in Mathematics from the University of Arizona.

Ray K. Miller
Northwest Evaluation Association
30633 11th Ave. S.
Federal Way, WA 98003
Mr. Miller is the Executive Secretary for the NWEA and an educational consultant for research and evaluation and educational measurement. He received his MEd in School Psychological Services from the University of Oregon and is a licensed psychologist. In the past, Mr. Miller was the Director of the Assessment and Evaluation Cooperative at ESD 121 in Washington and a part-time instructor at Seattle Pacific University and Central Washington University.

Connie Missimer
4836 N.E. 40th Street
Seattle, WA 98105
Ms. Missimer is author of *Good Arguments: an Introduction to Critical Thinking* (Prentice-Hall, 1986). For 8 years she taught high school and has recently begun work with primary grades. She has given numerous workshops with high school faculties. Last spring, she offered an assessment of critical thinking assessment tests to the Educational Testing Service in Berkeley.

Richard Naccarato, PhD
Northwest Regional Educational Laboratory
101 S.W. Main Street, Suite 500
Portland, OR 97204
After having taught junior and senior high school in Seattle, Dr. Naccarato obtained his PhD in educational psychology from the University of Washington. Since then he has coordinated placement and proficiency testing and taught measurement courses at the University of Iowa, worked for Educational Testing Service and consulted in testing and evaluation in several foreign countries. He currently is a research associate with the Test Center at Northwest Regional Educational Laboratory in Portland.

Wayne Neuberger, PhD
Director, Assessment and Evaluation
Oregon Department of Education
700 Pringle Parkway S.E.
Salem, OR 97310
Before assuming his current responsibilities with the State Department of Education, Dr. Neuberger was the Coordinator of Program Planning and Evaluation for Beaverton Schools. He is also a NWEA Board Member. Dr. Neuberger earned his PhD in Educational Research from New Mexico University.

Stephen P. Norris, PhD
Center for the Study of Reading
174 Children's Research Center
51 Gerty Drive
Champaign, IL 61820
Dr. Norris is a professor at the Memorial University in New Foundland and the author of the *Test on Appraising Observations*. He is currently a visiting scholar at the University of Illinois-Urbana Center for the Study of Reading. He is working on several other tests of higher order thinking and is coauthor with Robert Ennis of *Evaluating Critical Thinking* (to be released this summer). Dr. Norris earned his PhD from the University of Illinois-Urbana.

Lynde S. Paule, PhD
Northwest Regional Educational Laboratory
101 S.W. Main Street, Suite 500
Portland, OR 97204
Dr. Paule has worked in Evaluation and Assessment at NWREL for the past ten years. She developed test items for the states of Alaska, Washington and Hawaii and most recently, developed a Communications Test for Lewiston, Idaho. She has six years of classroom teaching experience and is currently an Instructor in the Management Division at Marylhurst College. She earned her PhD from the University of Oregon.

Edys S. Quellmalz, PhD
205 Georgia Lane
Portola, CA 94025
Dr. Quellmalz is a consultant to several education agencies' higher order thinking projects. She is an author of many books and articles on the use of thinking skills across the curriculum, including *Measuring Thinking Skills in the Classroom* which she coauthored with Richard Stiggins. Dr. Quellmalz was formerly a Senior Research Associate with the Center for the Study of Evaluation, UCLA. She earned her PhD in Instructional Research and Development from UCLA.

Stuart Rankin, EdD
Deputy Superintendent
Detroit Public Schools
5057 Woodward
Detroit, MI 48202
Dr. Rankin is Deputy Superintendent of Detroit Public Schools. He brings to this conference a perspective which combines curriculum, research and school administration. Dr. Rankin is coauthoring a book entitled *Dimensions in Thinking*. He has also been involved in a nationwide panel examining thinking skills. Dr. Rankin earned his EdD at Wayne State University.

Gale H. Roid, PhD
Assessment Research
P.O. Box 8900-324
Salem, OR 97303-0890
Dr. Roid has a PhD in psychometrics, has been a staff member and consultant to test publishers, has worked in statewide assessment, is a researcher with numerous publications in criterion-referenced testing, and is author of *A Technology of Test-item Writing* (Academic Press, 1982). He has assisted Psychological Corporation with several of their achievement and aptitude tests and he will be representing them today on the Test Publishers' Panel.

Catherine Ross
15820 6th S.W.
Seattle, WA 98166
Ms. Ross is a consultant in the fields of curriculum development and higher order thinking skills. She teaches at Seattle Pacific University and is currently an administrator in the Highline School District in Seattle. She is the coauthor of the *Ross Test of Higher Cognitive Processes* and author of *Cognitive Challenge Cards*. Ms. Ross received her EdM at University of Washington.

Richard Stiggins, PhD
Northwest Regional Educational Laboratory
101 S.W. Main Street, Suite 500
Portland, OR 97204
Dr. Stiggins is Director of the Center for Performance Assessment at NWREL and has developed strategies for classroom assessment of thinking. He is coauthor with Quellmalz and Rubel of *Measuring Thinking Skills in the Classroom*. He earned his PhD from Michigan State University and is a member of the faculty in the Graduate School of Professional Studies at Lewis and Clark College.

Vida S. Taylor, EdD
ESD #112
1313 N.E. 134th Street
Vancouver, WA 98685
Dr. Taylor was recently hired as the Supervisor for Curriculum Services for ESD 112. Dr. Taylor previously was an Elementary Principal for the Redmond, Oregon School District and a part-time instructor at Portland State University. She has an EdD in Educational Administration.

Selma Wassermann, EdD
Faculty of Education
Simon Fraser University
Burnaby, British Columbia
V5A 1S6 CANADA
Dr. Wassermann teaches at Simon Fraser University and has been involved with thinking skills for two decades. She and Louis E. Raths coauthored the seminal text, *Teaching for Thinking: Theory and Application* (Charles Merrill, 1966) and after Rath's death, Dr. Wassermann was executive author of the extensively revised edition, *Teaching for Thinking: Theory, Strategies, and Activities for the Classroom* (Teachers College Press, 1986). Dr. Wassermann earned her EdD from New York University.

Sponsors

Northwest Regional Educational Laboratory

The Northwest Regional Educational Laboratory (NWREL) assists education, government, community agencies, business, and labor in improving quality and equality of educational programs and processes. NWREL operates the Test Center, a library of assessment instruments and information about testing for the Northwestern states and Region IV Chapter 1 programs. NWREL is an independent, nonprofit institution governed by a 33-member Board of Directors.

ESD 112, Vancouver, Washington

Educational Service District 112 serves 31 public school districts in six counties in southwest Washington. It is one of 9 regional ESDs in the state. In addition, the ESD provides services to private schools and two state schools, and it functions as a liaison with universities and area community and industrial organizations. Approximately 80 percent of all services provided by ESD 112 are local district cooperatives designed to meet the expressed needs of the district's staff and students.

Idaho Association of School Administrators

The Association is a professional organization servicing the needs of public school superintendents, elementary and secondary school principals, and special education administrators in the state of Idaho.

Northwest Evaluation Association

The Northwest Evaluation Association (NWEA) is a consortium of school districts in the northwest devoted to the general improvement of educational assessment and evaluation. In addition, NWEA has developed assessment products including test item banks for the basic skill areas of reading, language arts, and mathematics. The Science Curriculum and Assessment Project has developed a 7,000 item pool. These item banks are Rasch calibrated and have been pilot tested extensively.

Office of the Superintendent of Public Instruction, Washington

The OSPI provides leadership in the state of Washington for the enhancement of thinking skills instruction. Two statewide conferences as well as numerous workshops have added impetus to this statewide thrust. Thinking skills objectives are integrated into all content areas in state curriculum guides.

Oregon Department of Education

The state of Oregon places a heavy emphasis on thinking skills in its curriculum. This is reflected both in the way in which test results are analyzed and in the state assessment tests currently under development.

Appendix B

Conference Participants

DR. JUDITH A. ARTER
NORTHWEST REGIONAL
EDUCATIONAL LAB
101 S.W. MAIN ST.
PORTLAND, OR 97204

DR. WILLIAM AUTY
CORVALLIS, ORE. SCH. DIST. 509J
1555 S.W. 35TH STREET
CORVALLIS, OR 97333

MR. FRED BANNISTER
WASH. OFFICE/SUPT. OF PUBLIC
INST.
OLD CAPITOL BUILDING, FG-11
OLYMPIA, WA 98504

MR. ROGER BERG
ALOHA HIGH SCHOOL, ALOHA,
ORE.
P.O. BOX 200; SCH. DIST. #48
BEAVERTON, OR 97075

MR. TOM BESSON
SCIENCE RESEARCH ASSOCIATES,
INC.
2505 MADISON ST.
EUGENE, OR 97405

MR. ERIC BIGLER
CLACKAMAS ESD
P.O. BOX 216
MARYLHURST, OR 97036

DR. DIAN BLOM
KENT, WASH. PUBLIC SCHOOLS
12033 S.E. 256TH
KENT, WA 98031

MR. CHARLES BLONDINO
SEATTLE, WASH. ESD NO. 121
1410 S. 200TH STREET
SEATTLE, WA 98198

MS. SANDY BLONDINO
SUMNER SCHOOL DISTRICT #320
1202 WOOD AVENUE
SUMNER, WA 98390

MR. PAUL M. BOLY
SUNSET HIGH SCHOOL
RTE. 2 BOX 612
HILLSBORO, OR 97123

MR. RICHARD BRANNAN
PORTLAND PUBLIC SCHOOLS
3830 S.E. 14TH AVENUE
PORTLAND, OR 97202

DR. KENNETH BUMGARNER
BETHEL SCHOOL DISTRICT #403
516 EAST 176TH
SPANAWAY, WA 98387

MS. SUE CHADWICK
PASCO, WASH. ESD. NO. 123
124 S. 4TH AVENUE
PASCO, WA 99301

MS. NANCY CONGLETON
SUMNER SCHOOL DISTRICT #320
1202 WOOD AVENUE
SUMNER, WA 98390

MR. GEORGE CRISP
CLACKAMAS, ORE. HIGH SCHOOL
13801 S. E. WEBSTER ROAD
MILWAUKIE, OR 97267

MR. NELS DOELEMAN
ALOHA HIGH SCHOOL, ALOHA,
ORE.
P.O. BOX 200; SCH. DIST. #48
BEAVERTON, OR 97075

MS. ELLEN EDDY
SPANAWAY, WASH. BETHEL SCH.
DIST.
516 E. 176TH STREET
SPANAWAY, WA 98387

DR. GARY ESTES
NORTHWEST REGIONAL
EDUCATIONAL LAB
101 S.W. MAIN STREET
PORTLAND, OR 97204

MS. DOROTHY FANNING
KENNEWICK, WASH. SCHOOL
DISTRICT
10205 MAPLE LANE
PASCO, WA 99301

MR. GLEN FIELDING
OREGON STATE BOARD OF EDUCATION
345 NORTH MONMOUTH AVENUE
MONMOUTH, OR 97361

MR. LANDON A. FITCH
WENATCHEE, WASH. SCHOOL DISTRICT
1101 MILLERDALE
WENTACHEE, WA 98801

MS. JILL FOLTZ
GRANTS PASS, ORE
N. VALLEY H. S.
6741 MONUMENT DRIVE
GRANTS PASS, OR 97526

MR. DAVID GODFREY
MILWAUKIE, ORE. HIGH SCHOOL
11300 S. E. 23RD
MILWAUKIE, OR 97222

MS. MARCIA HAACK
BEAVERTON, ORE. SCHOOL DISTRICT
P.O. BOX 200
BEAVERTON, OR 97075

MR. MARK HADDOCK
OLYMPIA, WASH. PUBLIC SCHOOLS
1113 E. LEGION WAY
OLYMPIA, WA 98501

MS. ARCELLA HALL
PORT ANGELES, WASH. HIGH SCHOOL
304 E. PARK AVENUE
PORT ANGELES, WA 98362

MS. ANITA HALSTEAD
CREATIVITY NETWORK
310 UNION AVENUE
SNOHOMISH, WA 98290

MS. MARILYN S. HARTZELL
NORTHWEST REGIONAL EDUCATIONAL LAB.
101 SW MAIN, SUITE 500
PORTLAND, OR 97204

DR. WALTER HATHAWAY
PORTLAND PUBLIC SCHOOLS
501 N. DIXON STREET
PORTLAND, OR 97227

MS. BEV HENDERSON
PASCO, WASHINGTON ESD 123
124 S. 4TH
PASCO, WA 99301

DR. MARSHALL HERRON
CORVALLIS, ORE. SCH. DIST. 509J
1555 S.W. 35TH STREET
CORVALLIS, OR 97333

MS. MARGARET HILL
SCIENCE RESEARCH ASSOCIATES, INC.
155 NORTH WACKER DR.
CHICAGO, IL 60606

DR. CAROLYN S. HUGHES
OKLAHOMA CITY PUBLIC SCHOOLS
900 N. KLEIN
OKLAHOMA CITY, OK 73106

MS. LINDA HUGLE
JOSEPHINE CTY. ORE. SCHOOL DISTRICT
706 N.W. A
GRANTS PASS, OR 97526

MS. MELINDA IRWIN
CLACKAMAS, ORE. HIGH SCHOOL
13801 S. E. WEBSTER ROAD
MILWAUKIE, OR 97267

DR. MARY M. JOHNSON
CORVALLIS, ORE. SCHOOL DIST. 509J
1555 SW 35TH
CORVALLIS, OR 97330

MS. PEGGY JOHNSON
LAKE STEVENS, WASH. SCHOOL DISTRICT
12708 20TH STREET N.E.
LAKE STEVENS, WA 98258

MS. BARBARA JOHNSTONE
WASH. OFF. OF SUPT. OF PUBLIC INST.
OLD CAPITOL BUILDING, FG-11
OLYMPIA, WA 98504

MS. PAM KAISER
LAKE STEVENS, WASH. SCHOOL DISTRICT
12708 20TH STREET N.E.
LAKE STEVENS, WA 98258

MS. JUDY KING
SUMNER SCHOOL DISTRICT #320
1202 WOOD AVENUE
SUMNER, WA 98390

DR. GAGE KINGSBURY
PORTLAND, ORE. PUBLIC SCHOOLS
501 N. DIXON SE
PORTLAND, OR 97227

MS. JEAN KIRSCH
PORTLAND, ORE. MULTNOMAH CTY. ESD
220 S. E. 102ND
PORTLAND, OR 97216

MR. ROB LARSON
OREGON MUSEUM OF SCIENCE & INDUSTRY
4015 S. W. CANYON
PORTLAND, OR 97221

DR. RICHARD LAUGHLIN
CLATSOP CTY. ORE. ESD SUPT.
3199 MARINE DRIVE
ASTORIA, OR 97103

MS. JANET L LAUGHLIN
ASTORIA, ORE. EDUCATIONAL SERVICES
RTE. 1, BOX 913E
ASTORIA, OR 97103

MR. JAY LEET
TIGARD SCHOOL DISTRICT 23J
13137 S. W. PACIFIC HIGHWAY
TIGARD, OR 97223

DR. THOMAS S. LINDERSMITH
LAKE OSWEGO SCHOOL DIST. 7J
P.O. BOX 70
LAKE OSWEGO, OR 97034

MR. TOM LOWE
WASHOUGAL, WASH. HIGH SCHOOL
1201 39TH STREET
WASHOUGAL, WA 98671

MR. JOHN MADISON
BEAVERTON, ORE. HIGH SCHOOL
P.O. BOX 200
BEAVERTON, OR 97223

MS. E. ANN MARCHBANK
TYEE, WASH. HIGH SCHOOL
4424 SOUTH 188TH
SEATTLE, WA 98188

MS. ELAINE MATTSON
BEAVERTON, ORE. SCHOOL DISTRICT
P.O. BOX 200
BEAVERTON, OR 97075

MS. SUZY MCCAUSLAND
NORTHWEST REGIONAL EDUCATIONAL LAB
101 S.W. MAIN STREET
PORTLAND, OR 97204

MS. DONNA MCDONOUGH
CLACKAMAS, ORE. COMMUNITY COLLEGE
P. O. BOX 477
BORING, OR 97009

MR. JOHN MCMAHAN
PORTLAND, ORE. PUBLIC SCHOOLS
5700 N. E. 39TH
PORTLAND, OR 97211

MR. RICHARD MEINHARD
OREGON DEPARTMENT OF EDUCATION
700 PRINGLE PARKWAY SE
SALEM, OR 97310

MR. RAY K. MILLER
NORTHWEST EVALUATION ASSSOCIATION
30633 11TH AVE. S.
FEDERAL WAY, WA 98003

MS. CONNIE MISSIMER
4836 N.E. 40TH STREET
SEATTLE, WA 98105

MS. STEPHANIE MITCHELL
PORTLAND PUBLIC SCHOOLS
501 N. DIXON
PORTLAND, OR 97227

MS. CAROLYN MOILANEN
PORTLAND, ORE. PUBLIC SCHOOLS
501 N. DIXON SE
PORTLAND, OR 97227

MS. JANE MONROE
CLACKAMAS, ORE. HIGH SCHOOL
13801 S. E. WEBSTER ROAD
MILWAUKIE, OR 97267

MR. DAVID G. MOORE
ALBANY, ORE. LINN-BENTON ESD
905 4TH AVENUE SE
ALBANY, OR 97321

DR. RICHARD NACCARATO
NORTHWEST REGIONAL
EDUCATIONAL LAB
101 S.W. MAIN ST.
PORTLAND, OR 97204

MR. TOM NEAL
THE DALLES, ORE. WILSON ELEM.
SCH.
1413 W. 12TH
THE DALLES, OR 97058

MS. KAREN NELSON
BEAVERTON SCHOOL DISTRICT
P.O. BOX 200
BEAVERTON, OR 97075

DR. WAYNE NEUBERGER
OREGON DEPARTMENT OF
EDUCATION
700 PRINGLE DRIVE
SALEM, OR 97310

MR. FRED E. NEWTON
PORTLAND, ORE. MULTNOMAH
CTY. ESD
220 SE 102ND STREET
PORTLAND, OR 97216

MS. SHARON NORLEN
CLACKAMAS, ORE. HIGH SCHOOL
13801 S. E. WEBSTER ROAD
MILWAUKIE, OR 97267

DR. STEPHEN P. NORRIS
CENTER FOR THE STUDY OF
READING
51 GERTY DRIVE
CHAMPAIGN, IL 61820

MS. DORIS O'HARA
CLACKAMAS, ORE. HIGH SCHOOL
13801 S. E. WEBSTER ROAD
MILWAUKIE, OR 97267

MS. DOROTHY PATRICK
PORTLAND, ORE. CENTENNIAL
SCH. DIST
18135 S. E. BROOKLYN STREET
PORTLAND, OR 97236

DR. LYNDE S. PAULE
NORTHWEST REGIONAL
EDUCATIONAL LAB
101 S.W. MAIN ST.
PORTLAND, OR 97204

MR. LEON PAULSON
MULTNOMAH CTY., ORE. ESD
220 S. E. 102ND
PORTLAND, OR 97216

MS. SANDRA PELLENS
OREGON DEPARTMENT OF
EDUCATION
700 PRINGLE PARKWAY S.E.
SALEM, OR 97370

MS. EILEEN PETERSEN
LAKE STEVENS, WASH. SCHOOL
DISTRICT
12708 20TH N.E.
LAKE STEVENS, WA 98258

DR. EDYS S. QUELLMALZ
205 GEORGIA LANE
PORTOLA, CA 94025

DR. STUART RANKIN
DETROIT PUBLIC SCHOOLS
5057 WOODWARD
DETROIT, MI 48202

MR. GORDON T. REES
SANDY, UT HIGHLAND HIGH
SCHOOL
11433 MT. RIDGE CIRCLE
SANDY, UT 84092

DR. GALE H. ROID
P.O. BOX 8900-324
SALEM, OR 97303

MS. CATHERINE ROSS
15820 6TH S.W.
SEATTLE, WA 98166

MS. JENNIFER R. SALMON
NORTHWEST REGIONAL
EDUCATIONAL LAB
101 S.W. MAIN ST.
PORTLAND, OR 97204

MR. STEPHEN SLATER
OREGON DEPARTMENT OF
EDUCATION
700 PRINGLE PARKWAY SE
SALEM, OR 97310

MS. MARGE SMITH
BEAVERTON, ORE. SCHOOL
DISTRICT
P. O. BOX 200
BEAVERTON, OR 97075

MS. MARLENE SODERQUIST
BEAVERTON, ORE. SCHOOL
DISTRICT
P.O. BOX 200
BEAVERTON, OR 97075

MS. JILL SORENSEN
SANDY, UTAH HIGHLAND HIGH
SCHOOL
709 DOUGLAS STREET
SALT LAKE CITY, UT 84102

DR. RICHARD STIGGINS
NORTHWEST REGIONAL
EDUCATIONAL LAB
101 S.W. MAIN ST.
PORTLAND, OR 97204

MR. TOM STRAUGH
ANCHORAGE, AK. SCHOOL
DISTRICT
4600 DEBARR, P.O.BOX 196614
ANCHORAGE, AK 99519

MS. SUE SWITA
KENNEWICK, WASH. SCHOOL
DISTRICT
1208 W. 19TH AVENUE
KENNEWICK, WA 99336

MS. SUE TARRANT-BERG
ALOHA HIGH SCHOOL
P.O. BOX 200; SCH. DIST. #48
BEAVERTON, OR 97075

DR. VIDA S. TAYLOR
ESD. NO. 112
1313 N.E. 134TH STREET
VANCOUVER, WA 98685

MR. JIM THOMAS
LAKE OSWEGO, ORE. SCHOOL
DISTRICT
18477 S.W. TIMBERGROVE COURT
LAKE OSWEGO, OR 97035

MS. JAN TIETZ
HIGHLINE PUBLIC SCHOOLS
15675 AMBAUM BLVD., SW
SEATTLE, WA 98166

MS. NANCY VAN KANNEL
TIGARD, OREGON HIGH SCHOOL
9000 JS.W. DURHAM ROAD
TIGARD, OR 97226

MS. DIANE WALKER
BETHEL SCH. DIST.
516 E. 176TH STREET
SPANAWAY, WA 98387

DR. JULIA C. WAN
BAINBRIDGE ISL. SCHOOL DIST.
8489 MADISON AVENUE NE
BAINBRIDGE ISLAND, WA 98110

DR. SELMA WASSERMANN
SIMON FRASER UNIVERSITY
BURNABY, BC V5A 1S6

MS. ANNE S. WAX
TIGARD SCHOOL DISTRICT 23J
13137 S. W. PACIFIC HIGHWAY
TIGARD, OR 97223

MS. RUTH WEHLING
1012 N. JANTZEN
PORTLAND, OR 97217

MR. KEN WELLS
CLACKAMAS, ORE. HIGH SCHOOL
13801 S. E. WEBSTER ROAD
MILWAUKIE, OR 97267

MR. E. PAUL WILLIAMS
PORTLAND, ORE. COMMUNITY
COLLEGE
12000 S. W. 49TH AVENUE
PORTLAND, OR 97219

DR. BARBARA WOLFE
OREGON DEPARTMENT OF
EDUCATION
700 PRINGLE PARKWAY SE
SALEM, OR 97310

MR. PETER WOLMUT
PORTLAND ORE., MULTNOMAH
CTY. ESD
P.O. BOX 16657
PORTLAND, OR 97211

MS. ILA ZBARASCHUK
AUBURN, WASH. ADVENTIST
ACADEMY
5000 AUBURN WAY SOUTH
AUBURN, WA 98002

MR. ROGER ZIEGELMAN
PSYCHOLOGICAL CORPORATION
2114-7TH AVE. W.
SEATTLE, WA 98119